Chapter 4: An outlook to dense graphs

Compilation date: December 19, 2017

1 Introduction

We have studied various properties of nowhere dense graph classes, most of which are naturally closed under taking subgraphs in the following sense. If a graph G has a property, say it excludes K_t as a depth-r minor for some values of t and r, then also every subgraph $H \subseteq G$ satisfies this property. When studying uniform quasi-wideness, we have in a sense forced this preservation property by stating a condition on every vertex subset $A \subseteq V(G)$. In this chapter we are going to study two properties of graphs that are not preserved under taking subgraphs and which still lead to a rich structural and algorithmic theory, namely the model theoretic notions of *stability* and the *non-independece property (NIP)*.

Nowhere dense graph classes are both stable and have the non-independence property, and, quite surprisingly, if we consider a class $\mathcal{C}$ of graphs which is closed under taking subgraphs and either stable or NIP, we find that $\mathcal{C}$ is nowhere dense. Stability theory and the theory of structures with the non-independence property will give us several powerful algorithmic tools, which are particularly useful in combination with methods from sparsity theory.

As model theory is not the focus of this course, we are going to give set theoretic definitions of the new concepts. Stability then corresponds to the *order dimension* of a set family and the non-independence property corresponds to the well known *Vapnik Chervonenkis dimension*, short *VC dimension*, of a set family.

2 Stability

The notion of stability in model theory was initially defined in terms of the number of first-order types over parameter sets, which can be seen as a generalization of distance-r neighborhood complexity as we have studied it in previous chapters. We are going to work with a more intuitive notion, which aims to measure the complexity of set systems in terms of forbidden structures in them.

2.1 Order dimension

Let A be a set and let $\mathcal{F}$ be a family of subsets of A; we will also call $\mathcal{F}$ a *set system* over the *ground set* A. With such a set system $\mathcal{F}$ we may associate its *incidence graph* $I(\mathcal{F})$, which is a bipartite graph with one side consisting of vertices of A and second side consisting of sets from $\mathcal{F}$, where $e \in A$ and $X \in \mathcal{F}$ are considered adjacent whenever $e \in \mathcal{F}$.

Definition 1. The *order dimension* of a set system $\mathcal{F}$ over A is the largest ℓ for which we may find elements $e_1, \ldots, e_\ell \in A$ and sets $X_1, \ldots, X_\ell \in \mathcal{F}$ such that $e_i \in X_j$ if and only if $i \leqslant j$, for all $i, j \in \{1, \ldots, \ell\}$. Elements $e_1, \ldots, e_\ell \in A$ and sets $X_1, \ldots, X_\ell \in \mathcal{F}$ having the property stated above are called a *ladder* of length ℓ.

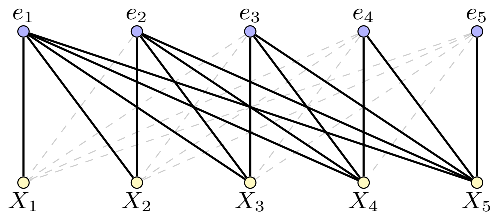

Figure 1: Ladder of length 5. Solid edges represent edges in $I(G)$, dashed grey edges represent non-edges in $I(G)$.

Thus, the order dimension of $\mathcal{F}$ is an upper bound on the length of ladders in the incidence graph $I(\mathcal{F})$, as depicted in Figure 1. The next proposition shows that the order dimension can be also interpreted in terms of lengths of chains in (restrictions) of $\mathcal{F}$. A *chain* in a set system $\mathcal{G}$ is any family of sets in $\mathcal{G}$ that is totally ordered by inclusion; the *length* of a chain is its cardinality. For a subset $B \subseteq A$ of the ground set, by $\mathcal{F}[B]$ we denote the *restriction* of $\mathcal{F}$ to B, that is,

$$\mathcal{F}[B] \coloneqq \{X \cap B \colon X \in \mathcal{F}\}.$$

Proposition 1. *The order dimension of a set system $\mathcal{F}$ over a finite ground set A is equal to the largest ℓ such that for all $B \subseteq A$, the chains in $\mathcal{F}[B] - \{\emptyset\}$ have length at most ℓ.*

Proof. On one hand, if we find a ladder of length ℓ consisting elements $e_1, \ldots, e_\ell \in A$ and sets $X_1, \ldots, X_\ell \in \mathcal{F}$, then setting $B \coloneqq \{e_1, \ldots, e_\ell\}$ we find that $\mathcal{F}[B] - \{\emptyset\}$ contains a chain of length ℓ, namely $X_1 \cap B, \ldots, X_\ell \cap B$. On the other hand, if for some B we find a chain $X_1 \cap B \subsetneq \ldots \subsetneq X_\ell \cap B$ in $\mathcal{F}[B] - \{\emptyset\}$, then by picking any elements $e_1 \in X_1 \cap B$ and $e_i \in (X_i \cap B) - (X_{i-1} \cap B)$ for $i > 1$ we obtain a ladder of length ℓ. □

Let us now give some examples to make these definitions more concrete. With any graph G and $r \in \mathbb{N}$, we may associate the set system of closed r-neighborhoods in G:

$$\mathrm{Balls}_r(G) = \{N_r[v] \colon v \in V(G)\}.$$

Thus, for an n-vertex graph, $\mathrm{Balls}_r(G)$ is a set system of size n over a ground set $V(G)$ of size n. Unraveling the definition of ladders an order dimension, the order dimension of $\mathrm{Balls}_r(G)$ is the largest ℓ for which we may find vertices $v_1, \ldots, v_\ell$ and $w_1, \ldots, w_\ell$ in G such that $\mathrm{dist}(v_i, w_j) \leqslant r$ if and only if $i \leqslant j$, for all $i, j \in \{1, \ldots, \ell\}$.

The following result was essentially proved during the tutorials, but we repeat it, as it provides the key connection between stability and sparsity.

Theorem 2. *Let $\mathcal{C}$ be a nowhere dense class of graphs. Then for every $r \in \mathbb{N}$ there exists an integer k such that for every graph $G \in \mathcal{C}$ the set family $\mathrm{Balls}_r(G)$ has order dimension at most k.*

Proof. Since $\mathcal{C}$ is nowhere dense, it is also uniformly quasi-wide, and this is witnessed by some functions $s \colon \mathbb{N} \to \mathbb{N}$ and $N \colon \mathbb{N} \times \mathbb{N} \to \mathbb{N}$ witness. Fix $r \in \mathbb{N}$ and a graph $G \in \mathcal{C}$. Let k be the order dimension of the set family $\mathrm{Balls}_r(G)$. That is, we may find vertices $v_1, \ldots, v_k$ and $w_1, \ldots, w_k$ of G such that $\mathrm{dist}_G(v_i, w_j) \leqslant r$ if and only if $i \leqslant j$, for all $i, j \in \{1, \ldots, k\}$.

Let $s = s(2r)$ and $t = 2 \cdot (r+2)^{s(2r)}$. We shall prove that $k < N(2r, t+1)$. For the sake of contradiction suppose this is not the case, that is, $k \geqslant N(2r, t+1)$. Let $W = \{w_1, \ldots, w_k\}$; then $|W| = k \geqslant N(2r, t+1)$. By uniform quasi-wideness, we can find disjoint vertex subsets $S \subseteq V(G)$ and $A \subseteq W - S$ such that $|S| \leqslant s$, $|A| > t$, and A is $2r$-independent in $G - S$.

For a vertex $w \in V(G)$, let $\pi_r[w, S] \colon S \to \{0, 1, \ldots, r, +\infty\}$ be its r-distance profile on S. Note that there are at most $(r+2)^s$ possible r-distance profiles on S. Since $|A| > t = 2 \cdot (r+2)^s$, we can find three indices $1 \leq \alpha < \beta < \gamma \leqslant k$ such that the vertices $a \coloneqq w_\alpha, b \coloneqq w_\beta, c \coloneqq w_\gamma$ belong to A and have equal r-distance profiles. Denote $d \coloneqq v_{\alpha+1}$. In particular, we have the following assertions

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

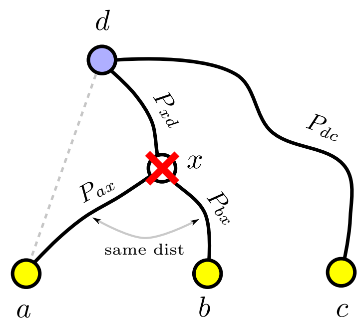

Figure 2: Proof of Theorem 2: contradiction looms.

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

This motivates the following definition.

Definition 2. A graph class $\mathcal{C}$ has *bounded order dimension* if there exists a function $f \colon \mathbb{N} \to \mathbb{N}$ such that for every $r \in \mathbb{N}$, the order dimension of $\mathrm{Balls}_r(G)$ for any $G \in \mathcal{C}$ is bounded by $f(r)$.

Corollary 3. *Every nowhere dense class has bounded order dimension.*

Let us see an example of a dense graph class with bounded order dimension. For a graph G and integer $s \in \mathbb{N}$, the *s-power* of G is the graph G^s on the same vertex set as G, where u, v are considered adjacent in G^s if and only if $\mathrm{dist}_G(u, v) \leqslant s$. Note that the power of a sparse graph is not necessarily sparse, e.g. the square of a star is a clique. The following observation follows immediately from the definitions.

Proposition 4. *If G is a graph and $r, s \in \mathbb{N}$, then* $\mathrm{Balls}_r(G^s) = \mathrm{Balls}_{rs}(G)$.

Corollary 5. *For every nowhere dense class $\mathcal{C}$ and $s \in \mathbb{N}$, the class $\mathcal{C}^s := \{G^s : G \in \mathcal{C}\}$ has bounded order dimension.*

Proof. By Proposition 4, if function $f(r)$ witnesses that $\mathcal{C}$ has bounded order dimension, then function $g(r) := f(rs)$ witnesses that $\mathcal{C}^s$ has bounded order dimension. □

2.2 Stability

In model theoretic terms, one would say that the property $\mathrm{dist}(x, y) \leqslant r$ (which can be expressed as a first-order formula) has the *k-order property* if there is no ladder in $\mathrm{Balls}_r(G)$ longer than k. More generally, for a first-order formula $\varphi(\bar{x}, \bar{y})$ with its free variables partitioned into two tuples $\bar{x}$ and $\bar{y}$ is said to have the k-order property in a graph G one can find tuples $\bar{a}_1, \ldots, \bar{a}_k$ and $\bar{b}_1, \ldots, \bar{b}_k$ of vertices of G, with $|\bar{a}_i| = |\bar{x}|$ and $|\bar{b}_i| = |\bar{y}|$ for all $1 \leqslant i \leqslant k$, such that

$$G \models \varphi(\bar{a}_i, \bar{b}_j) \;\Leftrightarrow\; i \leqslant j$$

for all $i, j \in \{1, \ldots, k\}$. A class $\mathcal{C}$ of graphs is called *stable* if for every formula $\varphi(\bar{x}, \bar{y})$ there exists an integer k such that φ does not have the k-order property on all graphs from $\mathcal{C}$. Thus, our notion of bounded order dimension is a relaxation of stability, where we consider only formulas $\delta_r(x, y)$ stating that $\mathrm{dist}(x, y) \leqslant r$, for all $r \in \mathbb{N}$. We remark that the above notion is usually considered for arbitrary relational structures, instead of just graphs.

We are not going to follow this path but only state the following fundamental fact, which says that in a fixed nowhere dense class $\mathcal{C}$, every first order formula $\varphi(\bar{x}, \bar{y})$ does not have the k-order property for some finite k.

Theorem 6. *Every nowhere dense class is stable.*

2.3 The Erdős-Hajnal property

We will now see some interesting application of the notion of bounded order dimension. 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 d 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 $\Omega(n^d)$. 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).

We will now show that in the presence of a bound on the order dimension of $\mathrm{Balls}_1(G)$, the Erdős-Hajnal conjecture holds. Note that the order dimension of $\mathrm{Balls}_1(G)$ is exactly the largest number ℓ for which we can find vertices $v_1, \ldots, v_\ell$ and $w_1, \ldots, w_\ell$ in G such that for all $i, j \in \{1, \ldots, \ell\}$, we have $v_i w_j \in E(G)$ or $v_i = w_j$ if and only if $i \leqslant j$.

Theorem 7. *Let G be a graph such that $|V(G)| > (a+b)^{2k+2}$ and $\mathrm{Balls}_1(G)$ has order dimension at most k. Then G contains either an clique of size a or an independent set of size b.*

We split the proof into several lemmas. 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 the order dimension of $\mathrm{Balls}_1(G)$ is at most 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 son, otherwise, replace w by its daughter. 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, $1 \leqslant \alpha \leqslant |w|$, 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 order dimension of $\mathrm{Balls}_1(G)$ gives us an upper bound on the alternation rank of τ.

Lemma 8. *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 $i \leqslant j$, for all $1 \leqslant i \leqslant \ell$. Since we assumed that $\mathrm{Balls}_1(G)$ has order dimension at most k, this implies that $\ell \leqslant k$, 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 9. *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

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

Corollary 10. *The tree τ has depth at least $a + b$.*

Proof. If τ had depth less than $a + b$, then by Lemmas 8 and 9, 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 Theorem 7. By Corollary 10, 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 S or at least b letters D. In the first case, let A be the set of all vertices $\tau(u)$ for which $u\mathrm{S}$ is a prefix of w. Then $|A| \geqslant a$ 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 Theorem 7.

By combining Theorem 2, Proposition 4, and Theorem 7, we immediately obtain the following.

Corollary 11. *Suppose $\mathcal{C}$ is a nowhere dense class of graphs. Then for every $r \in \mathbb{N}$ there exists an integer $t \in \mathbb{N}$ such that every n-vertex graph $G \in \mathcal{C}$ contains either a vertex subset A of size $\Omega(n^{\frac{1}{t}})$ whose all vertices are either pairwise at distance at most r or pairwise at distance more than r.*

3 Vapnik-Chervonenkis 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. Let $\mathcal{F}$ be a set system over a finite ground set A. We say that a set $B \subseteq A$ is *shattered* by $\mathcal{F}$ if $\mathcal{F}[B] = \mathcal{P}(B)$; that is, for every $C \subseteq B$ there exists $X \in \mathcal{F}$ such that $C = X \cap B$. We call the size of the largest set $B \subseteq A$ which is shattered by $\mathcal{F}$ the *Vapnik-Chervonenkis dimension*, or short *VC dimension*, of $\mathcal{F}$. We define the VC dimension of a graph G to be the VC dimension of the set system of closed 1-neighborhoods $\mathrm{Balls}_1(G)$.

The following observation follows directly from the definitions.

Proposition 12. *Let $\mathcal{F}$ be a set system of order dimension k, then $\mathcal{F}$ has VC dimension at most k.*

Thus, in graph classes of bounded order dimension we have that the VC dimensions of graphs, as well as their powers, are bounded.

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

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

Proof. We prove that $\mathcal{F}$ shatters at least $|\mathcal{F}|$ different subsets of the ground set A, which 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 less than $|\mathcal{F}|$. As $\mathcal{F}$ contains at least two sets, there exists $x \in A$ that belongs to some but not all of the sets in $\mathcal{F}$. We split $\mathcal{F}$ into two subsystems $\mathcal{X} = \{F \in \mathcal{F} \colon x \in F\}$ 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 $B \subseteq A$ is shattered by both set systems. Note that such a set B cannot contain x, since a set that contains x cannot be shattered by a system in which all sets contain x, or by a system in which all sets do not contain x. Hence, both B and $B \cup \{x\}$ are shattered by $\mathcal{F}$. This gives us for each set that is shattered both in $\mathcal{X}$ 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 $|\mathcal{X}| + |\mathcal{Y}| = |\mathcal{F}|$. □

We can also derive a polynomial bound on the neighborhood complexity of nowhere dense classes, analogous to Theorem 28 of Chapter 2.

Corollary 14. *For every nowhere dense class $\mathcal{C}$ and $r \in \mathbb{N}$, there exist a constant d, depending only on $\mathcal{C}$ and r, such that for every $G \in \mathcal{C}$ and $A \subseteq V(G)$, we have that*

$$|\{N_r^G[u] \cap A \colon u \in V(G)\}| \leqslant 1 + |A|^d.$$

Proof. Let $G \in \mathcal{C}$, $A \subseteq V(G)$ and $r \in \mathbb{N}$. According to Theorem 2 the family $\mathrm{Balls}_r(G)$ has order dimension at most d for some constant d, so it also has VC dimension at most d. It follows that the set system $\mathrm{Balls}_r(G)$ restricted to A also has VC dimension at most d, so by the Sauer-Shelah Lemma (Lemma 13) it follows that $|\mathrm{Balls}_r(G)[A]| \leqslant 1 + |A|^d$, which is equivalent to the claimed statement. □

Observe that Corollary 14 implies that even the number of r-neighborhood profiles on A is bounded polynomially. This is because an r-neighborhood profile of a vertex u on the set A is uniquely determined by $N_0^G[u] \cap A, N_1^G[u] \cap A, \ldots, N_r^G[u] \cap A$, and if the number of different i-neighborhoods $N_i^G[u] \cap A$ is bounded by $1 + |A|^{d_i}$, then the number of different r-neighborhood profiles on A is bounded by $\prod_{i=0}^{r}(1 + |A|^{d_i})$. In fact, similarly to bounded expansion classes, we have much better, almost linear bounds on the neighborhood complexity in nowhere dense classes, as made explicit in the following theorem.

Theorem 15. *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 r-neighborhood profiles on A is at most $c \cdot |A|^{1+\varepsilon}$.*

The proof of Theorem 15, which we will not present, follows exactly the same strategy as the proof for the bounded expansion case (Theorem 29 of Chapter 2). The difference is that some sets that are of constant size in the bounded expansion case, are of size at most $|A|^\varepsilon$ in the nowhere dense case. Applying a naive bound on the number of r-distance profiles on such sets would yield a disallowed exponential explosion in the bounds, but using the polynomial bounds given by Corollary 14, this explosion may be limited to polynomial, which is fine (ε gets rescaled to $d\varepsilon$). One additional ingredient that is needed is a bound on the VC dimension of the sets system of by weak r-reachability sets in any vertex ordering of the considered graph; we will take a closer look at them during the tutorials.

4 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 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 $\mathcal{F}$.

The greedy algorithm. 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 16. *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

$m_2 = m_1 - m_1/k \leqslant m_1 \cdot (1-1/k) \leqslant m \cdot (1-1/k)^2$. 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 $i \geqslant k \ln m$ 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 16 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.

4.1 Approximating hitting sets in set systemsm of bounded VC dimension

In this section we will prove the following theorem.

Theorem 17. *Let $d \geqslant 2$ 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 17. 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 4. 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 $\mu(F) \geqslant \epsilon$.

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 18. *Let $\mathcal{F}$ be a set system of VC dimension $d \geqslant 2$ and let μ be a probability distribution on the ground set A. There exists a universal constant c such that for every $0 < \epsilon \leqslant 1/2$, 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 19. *Let $X_1, \dots, X_n$ be i.i.d. Bernoulli random variables with success probability p. Let $X = X_1 + \dots + X_n$ and let $\mu = np$ be the expected value of X. Then for every $\delta \in [0,1]$, we have*

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

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

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

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

Proof of Theorem 18. 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 $\mu(F) \geqslant \epsilon$, 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 $(1-\epsilon)^s \leqslant e^{-s\epsilon}$. 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, $\mathbb{P}[E_2] \leqslant \mathbb{P}[E_1]$, since E_2 in particular requires E_1 to occur. We are first going to show that $\mathbb{P}[E_2] \geqslant \frac{1}{2}\mathbb{P}[E_1]$.

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 $\mathbb{P}[E_2 \mid N] \geqslant \mathbb{P}[|M \cap F_N| \geqslant k]$. 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 19 for $n = s$, $p = \epsilon$, and $\mu = s\epsilon = cd \ln \frac{1}{\epsilon}$, we have

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

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

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

[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

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

If $|S \cap F| < k$ then $p_F = 0$. Otherwise, we have $p_F \leqslant \mathbb{P}[N \cap F = \emptyset \mid S]$. 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

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

Finally we use that $\mathcal{F}$ has bounded VC dimension. According to the Sauer-Shelah lemma, Lemma 13, the number of distinct intersections of $\mathcal{F}$ with the sequence S is at most

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

here we used the fact that $\binom{n}{k} \leqslant \left(\frac{en}{k}\right)^k$ for all $1 \leqslant k \leqslant n$. Since the event that $N \cap F = \emptyset$ and $|M \cap F| \geqslant k$ 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

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

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

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

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

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

for such choice of c, implying

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

This finishes the proof of the theorem. □

Theorem 18 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:**
  - $\sum_{a\in F} \mu_a \geqslant \epsilon$ for all $F \in \mathcal{F}$;
  - $\mu_a \geqslant 0$ 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: $\mu^\star(F) \geqslant \epsilon^\star$ for each $F \in \mathcal{F}$. For now assume $\epsilon^\star \leqslant \frac{1}{2}$; we will treat the remaining corner case $\epsilon^\star > \frac{1}{2}$ later. Now we apply Theorem 18 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 $\mu^\star(F) \geqslant \epsilon^\star$ 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:**
  - $\sum_{a\in F} x_a \geqslant 1$ for all $F \in \mathcal{F}$; and
  - $x_a \geqslant 0$ 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 $\tau^\star(\mathcal{F}) \leqslant \tau(\mathcal{F})$. 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 20. *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 17: the sample size $c \cdot \frac{d}{\epsilon^\star} \ln \frac{1}{\epsilon^\star}$ is in fact bounded by

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

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

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

Finally, by applying Theorem 17 to the set system of r-balls in a graph from a nowhere dense class we infer the following.

Corollary 22. *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 an* r*-dominating set in* G *of size at most* $\mathcal{O}(k \log k)$*, where* k *is the minimum size of an* r*-dominating set in* G*.*

We remark that the algorithm of Theorem 17 can in fact be derandomized, yielding the same asymptotic bound on the approximation ratio. This is done by replacing Theorem 18 with a deterministic counterpart with the same asymptotic bounds.