VC-dimension, regularity, and the Erdős–Hajnal conjecture

Talk by Oriol Solé Pi

Notes by Sanjana Das

December 8, 2023

§1 VC-dimension

VC-dimension is a way to attempt to define dimension of purely combinatorial objects; we want to say that if the VC-dimension is low, then the object behaves like a low-dimensional geometric object in some way.

VC-dimension deals with objects of the form $(X, \mathcal{F})$ where X is some ground set and $\mathcal{F}$ is a set of subsets of X, called *ranges*.

Definition 1.1. For $S \subseteq X$, we define the *projection of $\mathcal{F}$ onto S*, denoted $\mathcal{F}|_S$, as $\{S \cap f \mid f \in \mathcal{F}\}$.

Definition 1.2. We say that a set $S \subseteq X$ is *shattered* if $|\mathcal{F}|_S| = 2^{|S|}$.

In other words, S is shattered if $\mathcal{F}|_S$ has all the possible sets you could get (i.e., all the subsets of S).

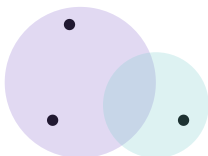
Definition 1.3. The *VC-dimension* of $(X, \mathcal{F})$ is the largest d for which there exists a set $S \subseteq X$ of size d which is shattered.

Example 1.4

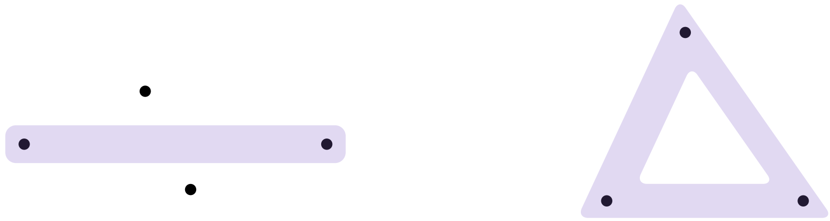
Suppose X is a set of points in the plane, and $\mathcal{F}$ is the set of subsets you can get by intersecting these points with a disk. Then the VC-dimension of $(X, \mathcal{F})$ is 3.

— if you have 3 points then any two can be intersected with a disk. Meanwhile, if you have 4 points, then there's some subset you can't get (in the convex case, two opposite points; in the concave case, the three points on the hull).

To see this, if we have 3 points, then we can intersect any one or two (or none, or all three) with a disk (so the subset consisting of these three points is shattered).



Meanwhile, if we have 4 points, then no matter how they are arranged, there is some subset that we can't get. If the four points are in convex position, then we can't get a disk that contains just one pair of opposite points; if they are not in convex position (so we have a triangle with one point inside it), we can't get a disk that contains just the three points forming the triangle.



We'll now look at some properties of the VC-dimension. We always use $(X, \mathcal{F})$ to denote our set system, and d to denote the dimension of $(X, \mathcal{F})$.

§1.1 The shatter lemma

Lemma 1.5 (Shatter lemma)
Let $(X, \mathcal{F})$ have dimension d. Then for any $Y \subseteq X$, we have $|\mathcal{F}|_Y| \leq \sum_{i=0}^{d} \binom{|Y|}{i} \leq |Y|^d$.

Proof. We'll prove that $|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{|X|}{i}$; the desired statement then follows from looking at the set system $(Y, \mathcal{F}|_Y)$ instead (which has dimension at most that of $(X, \mathcal{F})$, since any set $S \subseteq Y$ shatteered by $\mathcal{F}|_Y$ is also shattered by $\mathcal{F}$).

Given any $a \in X$, we can imagine shifting $\mathcal{F}$ downwards by a to obtain a new family $\mathcal{F}_a$ — this means we try to shift each $f \in \mathcal{F}$ downwards by a (i.e., remove a from f), and we do so if and only if this shifting would not cause a collision. More explicitly, for each $f \in \mathcal{F}$ we define

$$f_a = \begin{cases} f \setminus \{a\} & \text{if } f \setminus \{a\} \notin \mathcal{F} \\ f & \text{if } f \setminus \{a\} \in \mathcal{F}. \end{cases}$$

(If f does not contain a to begin with, then $f_a = f$.) This shifting operation doesn't affect the size of $\mathcal{F}$, since we ensure that no two sets collide when we shift them.

Claim 1.6 — We have $(X, \mathcal{F}_a)_{\text{dim}} \leq (X, \mathcal{F})_{\text{dim}}$ (where $(X, \mathcal{F})_{\text{dim}}$ denotes the dimension of $(X, \mathcal{F})$).

If we can prove this claim, then in order to finish, we can keep shifting $\mathcal{F}$ by elements $a \in X$ until we get stuck, in the sense that shifting doesn't allow us to change any of our sets. When that happens, the family $\mathcal{F}$ we have at that point must be downwards-closed, meaning that whenever $A \subseteq B$ and $B \in \mathcal{F}$ we have $A \in \mathcal{F}$ as well (otherwise, there would be some element we could shift by to change $\mathcal{F}$). At this point, all ranges in $\mathcal{F}$ must have size d — for every range $S \in \mathcal{F}$, all subsets of S are also in $\mathcal{F}$, so S is shattered, and since $(X, \mathcal{F})_{\text{dim}} \leq d$ (by the claim that shifting doesn't increase our dimension), no set of size at least $d + 1$ can be shattered. So at that point, we get that $|\mathcal{F}| \leq \sum_{i=0}^{d} \binom{|X|}{i}$ (and since shifting doesn't change the size of $\mathcal{F}$, the same is true of the original $\mathcal{F}$).

Now we'll prove the claim. To do so, it suffices to show that for any $S \subseteq X$ such that S is shattered in the new system $(X, \mathcal{F}_a)$, the same set S is also shattered in the original system $(X, \mathcal{F})$.

Consider any $B \subseteq S$; the fact that S is shattered means that there exists some $f_a \in \mathcal{F}_a$ — produced by shifting some $f \in \mathcal{F}$ — with $f_a \cap S = B$. If $a \notin S$ then there is nothing to prove (because $f \cap S = B$ as well — whether we include a in f or not doesn't affect its intersection with S). So we can assume that $a \in S$.

Case 1 ($a \in B$). Then f_a contains a, which means $f_a = f$ (since f_a is either obtained by removing a from f or by keeping f as it is, and in the first case it wouldn't contain a). So $f \cap S = B$ (where f is a range in the original family).

Case 2 ($a \notin B$). Then since S is shattered in $(X, \mathcal{F}_a)$ (so we can find an element of $\mathcal{F}_a$ achieving *any* intersection with S), there must be some other element $f'_a \in \mathcal{F}_a$ with $f'_a \cap S = B \cup \{a\}$. But then f'_a has to contain a, which means that f'_a wasn't shifted — i.e., it was obtained by taking some $f' \in \mathcal{F}$ and keeping it as it is. The only way this could happen is if removing a from f' would cause a collision — so then there has to be some $f'' \in \mathcal{F}$ with $f'' = f' \setminus \{a\}$. Then we have $f'' \cap S = B$, as desired. □

§1.2 Separated sets and the packing lemma

We would like to show that in some ways a set with low VC-dimension behaves like a geometric object of low dimension. In particular, in a low-dimensional space we can't have too many points which are far apart from each other. We'd like to prove a similar statement here — if we have some sets in $\mathcal{F}$ such that any two of them have large symmetric difference (which in some sense means they're far apart from each other), we'd like to show there aren't too many of them.

Theorem 1.7 (Packing lemma)

Let $(X, \mathcal{F})$ have dimension d and let $|X| = n$, and suppose $\mathcal{F}$ is such that $|S_1 \,\Delta\, S_2| \geq \delta$ for all distinct $S_1, S_2 \in \mathcal{F}$. Then $|\mathcal{F}| = O((n/\delta)^d)$.

Remark 1.8. This implies the slightly more general statement that if we have any $\{S_1, \ldots, S_k\} \subseteq \mathcal{F}$ such that $|S_i \,\Delta\, S_j| \geq \delta$ for all $i \neq j$, then $k = O((n/\delta)^d)$ — since we can simply replace $\mathcal{F}$ with the family $\{S_1, \ldots, S_k\}$ (removing sets from $\mathcal{F}$ can only decrease the VC-dimension). This is the version we'll actually use.

Remark 1.9. Here δ doesn't denote a quantity close to 0 — it's a constant greater than 1.

In the full version of the packing lemma, the hidden constant in $O((n/\delta)^d)$ is absolute; we will prove a weaker version, with a constant that depends on d.

To prove this, we need an intermediate result.

Definition 1.10. The *unit distance graph* for $(X, \mathcal{F})$ is the graph on vertex set $\mathcal{F}$ where two vertices are adjacent if their symmetric difference contains a single element.

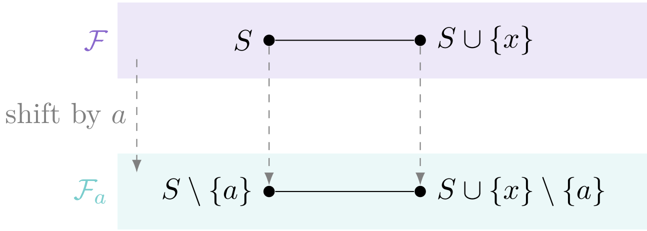
Lemma 1.11

The number of edges in the unit distance graph is at most $d\,|\mathcal{F}|$.

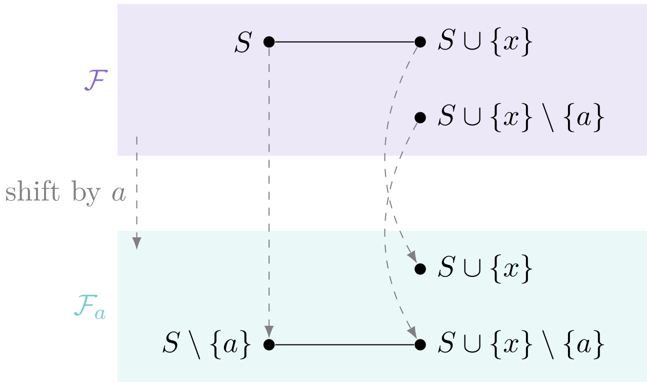
Proof. We'll again use the idea of shifting — we've seen earlier that shifting (i.e., replacing $\mathcal{F}$ with $\mathcal{F}_a$ for some $a \in X$) does not increase the dimension, and it preserves the number of ranges. We'll show that it also preserves the number of edges in the unit distance graph. Suppose we have some edge $S \sim S \cup \{x\}$ in the graph for $(X, \mathcal{F})$ (all edges are between two sets that differ only by one element). We'll show that when we replace $\mathcal{F}$ by $\mathcal{F}_a$, we'll either keep this edge, or get a new edge that compensates for it.

Case 1 ($a \notin S$). Then shifting by a doesn't affect either S or $S \cup \{x\}$ (this is certainly true if $a \notin S \cup \{x\}$, and if $x = a$ then it's still true because shifting $S \cup \{a\}$ would cause a collision with S), so there is nothing to prove (we keep the same edge $S \sim S \cup \{x\}$ in the new graph).

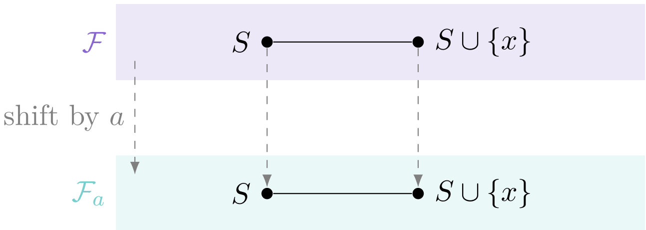
Case 2 ($a \in S$, and S gets shifted). If S gets shifted, then in $\mathcal{F}_a$ we now have $S \setminus \{a\}$ in its place. There are two possibilities. First, if $S \cup \{x\}$ *also* gets shifted (so it becomes $S \cup \{x\} \setminus \{a\}$), then we get the new edge $S \setminus \{a\} \sim S \cup \{x\} \setminus \{a\}$ between these two shifts.

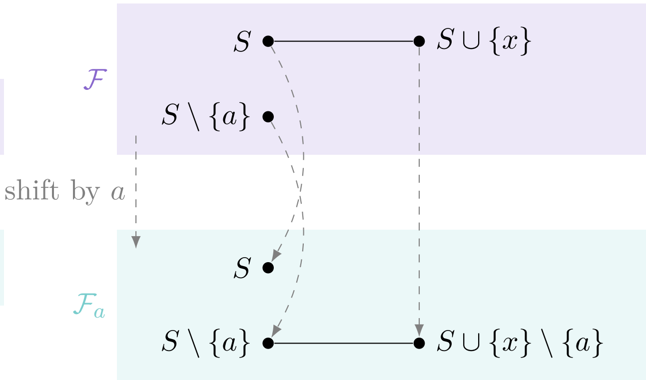


Meanwhile, if $S \cup \{x\}$ *doesn't* also get shifted (meaning that it remains as it is when we go from $\mathcal{F}$ to $\mathcal{F}_a$), then this means $S \cup \{x\} \setminus \{a\}$ was already in $\mathcal{F}$. So we still get a new edge $S \setminus \{a\} \sim S \cup \{x\} \setminus \{a\}$.



Case 3 ($a \in S$, but S doesn't get shifted). Then this means $S \setminus \{a\}$ was already in $\mathcal{F}$. If $S \cup \{x\}$ also doesn't get shifted, then we keep the same edge $S \sim S \cup \{x\}$ from before; and if $S \cup \{x\}$ does get shifted to $S \cup \{x\} \setminus \{a\}$, then we get a new edge $S \setminus \{a\} \sim S \cup \{x\} \setminus \{a\}$.





So we've shown that shifting doesn't affect the number of edges in the unit distance graph, which means we can again shift repeatedly until $\mathcal{F}$ is downwards closed. At this point we know every $f \in \mathcal{F}$ has size at most d. And now the inequality is easy to prove — for every edge e, 'assign' e to whichever of its two endpoints is larger. (So we assign the edge $f \sim f \cup \{x\}$ to the vertex $f \cup \{x\}$.) Then each element of $\mathcal{F}$ is assigned at most d edges (since the edges assigned to f all lead to sets f' obtained by deleting one element from f, and f only has d elements), so the number of edges is at most $d\,|\mathcal{F}|$. □

Proof of packing lemma. We'll start with $(X, \mathcal{F})$; then we'll take a uniform random subset $S \subseteq X$ of size $|S| = 4dn/\delta$ and project to it. On one hand, by the shatter lemma we know that $|\mathcal{F}|_S|$ is necessarily small; on the other hand, we'll prove an upper bound on $|\mathcal{F}|$ based on $\mathbb{E}\,|\mathcal{F}|_S|$, and combining these will give an upper bound on $|\mathcal{F}|$.

Choose $S \subseteq X$ of the appropriate size uniformly at random, and consider the set system $(S, \mathcal{F}|_S)$. We'll work with the unit distance graph for $(S, \mathcal{F}|_S)$, which has vertex set $\mathcal{F}|_S$; let its edge set be E (each edge connects elements of $\mathcal{F}|_S$ which differ in exactly one element). Since we really care about $\mathcal{F}$ and lots of elements of $\mathcal{F}$ could get mapped to the same element of $\mathcal{F}|_S$, we really want to work with a weighted version of this unit distance graph — define the weight of a *vertex* $i \in \mathcal{F}|_S$ as the number of sets $f \in \mathcal{F}$ which get projected to i (i.e., $w(i) = \#\{f \in \mathcal{F} \mid f \cap S = i\}$), and define the weight of an *edge* $e = i \sim j$ as $w(e) = \min\{w(i), w(j)\}$.

Remark 1.12. It might be more natural to instead define $w(e) = w(i)w(j)$. This definition works as well, but the one we use makes the computations a bit easier.

Claim 1.13 — We have $w(E) \leq 2d\,|\mathcal{F}|$.

(We use $w(E)$ to denote the total weight of all the edges in the graph.)

Proof. We can prove this by repeatedly applying the original (unweighted) inequality on the number of edges in a unit distance graph — we showed earlier that $|E| \leq d\,|\mathcal{F}|_S|$ (projecting to S can't increase the VC-dimension), so there must be some vertex $i \in \mathcal{F}|_S$ of degree at most $2d$. Then we can remove i from $\mathcal{F}|_S$ and see how much this decreases the weight in our graph — the weight of each edge incident to i is at most $w(i)$ (this is where the definition of edge weights as a minimum is convenient), so when we remove i, the total edge weight goes down by at most $2d \cdot w(i)$.

And now we can just keep repeating this — we repeatedly remove vertices until there are none left, keeping track of how much the total edge weight decreases at each step. (Note that removing elements of $\mathcal{F}|_S$ cannot increase the dimension of the system, so we can make the same argument (that there exists a vertex of degree $2d$) at each step.) In the end, this gives that the total edge weight is

$$w(E) \leq \sum_{i \in \mathcal{F}|_S} 2d \cdot w(i) = 2d \sum_{v \in \mathcal{F}|_S} w(i) = 2d\,|\mathcal{F}|$$

(since $w(i)$ is defined as the number of elements of $\mathcal{F}$ that project to i, so $\sum_v w(i) = |\mathcal{F}|$). □

This is one of two inequalities that will go into the proof of the bound, and the other is the following.

Claim 1.14 — We have $\mathbb{E}[w(E)] \geq 4d\,|\mathcal{F}| - 4d\mathbb{E}\,|\mathcal{F}|_S|$.

Proof. The idea is that instead of imagining that we sample all elements of S at the same time, we'll first sample the first $4dn/\delta - 1$ elements (uniformly at random, without repetition), and then separately sample the last element, which we'll call v. Every edge in E corresponds to some element of S (each edge connects two elements of $\mathcal{F}|_S$ which differ in exactly one element, and we can look at which element of S that one element is). We will just bound the (expected) total edge weight coming from edges that correspond to v (i.e., edges between two elements of $\mathcal{F}|_S$ with symmetric difference v); we let E_v denote the set of such edges. Then to find $\mathbb{E}[w(E)]$, by symmetry we can just multiply by the number of elements in $|S|$.

Let our original sample consisting of $4dn/\delta - 1$ elements be $Y \subseteq X$; we'll lower-bound the expected value of $w(E_v)$ conditioned on Y (the expectations below assume that we've already chosen Y, and are over the random choice of v).

We can think of E_v in the following way. Imagine we first start with $\mathcal{F}|_Y$, and then we go from Y to S by adding in v. Then some of the ranges in $\mathcal{F}|_Y$ are going to split into two ranges — if we have $q \in \mathcal{F}|_Y$ and q is a projection onto Y of ranges in $\mathcal{F}$ containing v as well as ranges in $\mathcal{F}$ *not* containing v, then it splits into both q and $q \cup \{v\}$ in $\mathcal{F}|_S$. And the edges in E_v are precisely those produced in this way (i.e., by considering some $q \in \mathcal{F}|_Y$ and taking the edge $q \sim q \cup \{v\}$ in $\mathcal{F}|_S$).

So let's look at some range $q \in \mathcal{F}|_Y$, and let $\mathcal{Q} \subseteq \mathcal{F}$ be the set of ranges in $\mathcal{F}$ that get projected to q. Let $b = |\mathcal{Q}|$; let b_v be the number of ranges in $\mathcal{Q}$ that contain v, and let $b'_v = b - b_v$ be the number of ranges in

$\mathcal{Q}$ that don't contain v. Then b_v and b'_v are the weights of $q \cup \{v\}$ and q in the unit distance graph for $\mathcal{F}|_S$, respectively, so the weight of the edge between them will be $\min\{b_v, b'_v\}$.

In order to show that $\mathbb{E}[w(E_v)]$ is large, we'll show that $\mathbb{E}[\min\{b_v, b'_v\}]$ is large for *each* individual $q \in \mathcal{F}|_Y$, and then sum the bound we get over all $q \in \mathcal{F}|_Y$.

To do so, it'll be easier to consider $\mathbb{E}[b_v b'_v]$ — which is the expected number of pairs of ranges $S_1, S_2 \in \mathcal{Q}$ whose projections become different when we add in v. (In $\mathcal{F}|_Y$, both S_1 and S_2 get projected to the same element q; we want to count pairs where when we add v, one gets projected to q and the other to $q \cup \{v\}$.)

Consider any two distinct ranges $S_1, S_2 \in \mathcal{Q}$; we are given that $S_1 \,\Delta\, S_2 \geq \delta$. Our original sample Y can't have chosen any elements of $S_1 \,\Delta\, S_2$ (since if it did, then S_1 and S_2 would not be projected to the same element of $\mathcal{F}|_Y$). Then S_1 and S_2 form such a pair (where they get projected to different elements in $\mathcal{F}|_S$) if and only if $v \in S_1 \,\Delta\, S_2$. And there are at least δ choices of v for which this occurs, out of at most n choices in total (there's actually even fewer choices than this, because we already chose Y), which means the probability this happens is at least

$$\mathbb{P}[S_1 \text{ and } S_2 \text{ end up with different projections}] \geq \frac{|S_1 \Delta S_2|}{n} \geq \frac{\delta}{n}.$$

So we've found the probability that each pair $S_1, S_2 \in \mathcal{Q}$ contributes to $b_v b'_v$ (the number of pairs for which this occurs); then by linearity of expectation we get

$$\mathbb{E}[b_v b'_v] \geq \frac{\delta}{n} \cdot \#\text{pairs } S_1, S_2 \in \mathcal{Q} = \frac{\delta}{n} \cdot |\mathcal{Q}| \, (|\mathcal{Q}| - 1).$$

Then since each of b_v and b'_v is at most $|\mathcal{Q}|$, we always have

$$\min\{b_v, b'_v\} \geq \frac{b_v b'_v}{|\mathcal{Q}|},$$

and therefore we have

$$\mathbb{E}[\min\{b_v, b'_v\}] \geq \frac{\delta}{n}(|\mathcal{Q}| - 1).$$

Finally, this is is the expected contribution to $w(E_v)$ coming from *one* range $q \in \mathcal{F}|_Y$ (and its corresponding $\mathcal{Q}$). So to bound $\mathbb{E}[w(E_v)]$, we can sum over all $q \in \mathcal{F}|_Y$. We have $\sum_q |\mathcal{Q}| = |\mathcal{F}|$ (since $\mathcal{Q}$ is the set of ranges in $\mathcal{F}$ projected to q in $\mathcal{F}|_Y$, and each range in $\mathcal{F}$ is projected to one range in $\mathcal{F}|_Y$), so summing over q gives

$$\mathbb{E}[w(E_v)] \geq \frac{\delta}{n}(|\mathcal{F}| - |\mathcal{F}|_Y|).$$

Finally, accounting for the random choice of Y as well, we have

$$\mathbb{E}[w(E_v)] \geq \frac{\delta}{n}(|\mathcal{F}| - \mathbb{E}\,|\mathcal{F}|_Y|).$$

This gives the expected amount of weight in the unit distance graph corresponding to *one* element of S, and since S has $4dn/\delta$ elements, by symmetry we have

$$\mathbb{E}[w(E)] = \frac{4dn}{\delta}\mathbb{E}[w(E_v)] \geq 4d(|\mathcal{F}| - \mathbb{E}\,|\mathcal{F}|_Y|) \geq 4d\,|\mathcal{F}| - 4d\mathbb{E}\,|\mathcal{F}|_S|$$

(since $|\mathcal{F}|_Y| \leq |\mathcal{F}|_S|$). □

Now combining the bounds from the two claims, we have $2d\,|\mathcal{F}| \geq w(E)$ (for all choices of S) and $\mathbb{E}[w(E)] \geq 4d\,|\mathcal{F}| - 4d\mathbb{E}\,|\mathcal{F}|_S|$ (over the random choice of S), so

$$|\mathcal{F}| \leq 2\mathbb{E}\,|\mathcal{F}|_S|\,.$$

But finally, we know that $|\mathcal{F}|_S| \leq |S|^d$ by the shatter lemma. So this gives us

$$|\mathcal{F}| \leq 2\,|S|^d = 2\left(\frac{4dn}{\delta}\right)^d = 2 \cdot (4d)^d \cdot \left(\frac{n}{\delta}\right)^d,$$

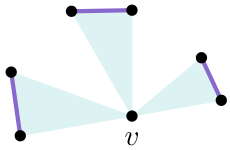
which is what we wanted to prove. (We have a factor of $(4d)^d$ that really doesn't need to be there — it's possible to get an absolute constant — but getting rid of it is tricky.) □

§2 Graphs and regularity

§2.1 VC-dimension for hypergraphs

Definition 2.1. Let $\mathcal{H}$ be a k-uniform hypergraph. For each vertex $v \in V$, we define its *neighborhood* $N(v)$ as the subset of $\binom{V}{k-1}$ consisting of all $(k-1)$-tuples such that if we add v to them (to form a k-tuple), we get an edge of $\mathcal{H}$.

In the 2-uniform case, the neighborhood of v is what you'd expect; in the 3-uniform case, it'll consist of pairs of vertices which, together with v, form edges.



(Here the edges of $\mathcal{H}$ are shown in blue, and the pairs in $N(v)$ in purple.)

Definition 2.2. We say the VC-dimension of a hypergraph $\mathcal{H}$ is the VC-dimension of the set system $(\binom{V}{k-1}, \mathcal{N})$, where $\binom{V}{k-1}$ is the set of all $(k-1)$-tuples of vertices and $\mathcal{N}$ is the set of all neighborhoods.

So here our ground set is all $(k-1)$-tuples of vertices, and each range is a neighborhood of some vertex.

§2.2 A regularity lemma

We'd like to say that graphs with low VC-dimension satisfy a very strong kind of regularity.

Lemma 2.3 (Super-strong regularity lemma)

If $\mathcal{H}$ is a k-uniform hypergraph with VC-dimension at most d, then for every $\varepsilon \in (0, \frac{1}{2})$, there exists an equipartition of the vertex set into $O(1/\varepsilon^{2d+1})$ parts such that for all but an ε-fraction of k-tuples of parts, the density between these k parts is at most ε or at least $1-\varepsilon$.

Compared to the usual regularity lemma, here we get a partition of size polynomial in ε^{-1}.

For illustration, imagine that $\mathcal{H}$ is 5-uniform; then this states that whenever we look at a 5-tuple of parts, for all but an ε-fraction of such 5-tuples, the density between them is very close to 0 or very close to 1. The density between a 5-tuple of parts is defined as

$$d(V_1, \dots V_5) = \frac{e(V_1, \dots, V_5)}{|V_1| \cdots |V_5|}$$

(as with usual graphs, we define density by taking the actual number of edges present, and dividing by the number of edges we'd have if the graph were complete).

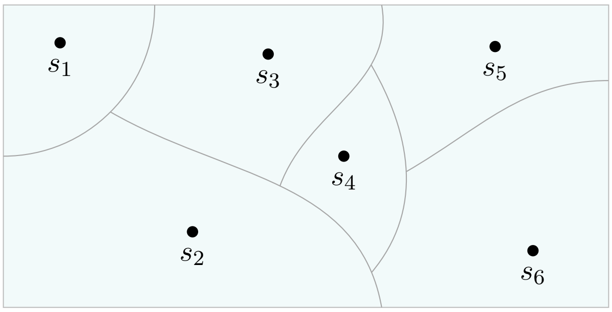
Counting lemmas for hypergraphs are in general a very delicate matter, but when you have this strong condition that all densities are very small or large, things become much easier because of the union bound. For example, suppose we have a 4-tuple of parts in a 3-uniform hypergraph, and we want to count the number of tetrahedra between them. In general, just having access to the densities between all triples of parts doesn't tell you the number of tetrahedra. But if we know that all the densities are at least $1-\varepsilon$, then if we choose a random vertex from each of the sets, they'll form a tetrahedron with probability at least $1-4\varepsilon$ by the union bound. Meanwhile, if one of the densities is at most ε, then we get a tetrahedron with probability at most ε. So this statement is quite useful.

> **Remark 2.4.** If you don't require that the partition is an equipartition, then you can get rid of the exceptional ε-fraction of k-tuples.

Proof sketch. We'll discuss the main ideas of the proof.

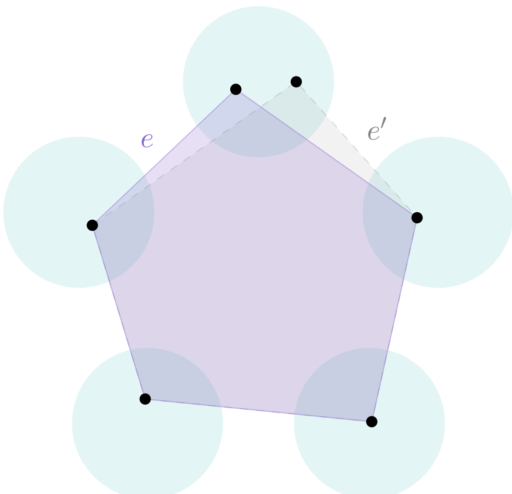
First, we look at the vertex set V of $\mathcal{H}$, and we greedily construct a set of vertices whose neighborhoods are very different from each other (in the sense of having large symmetric differences) — so we choose a set of vertices $S = \{s_1, \ldots, s_t\}$ such that $|N(s_i) \mathbin{\Delta} N(s_j)| \geq \delta$ for all $i \neq j$. (We keep choosing vertices greedily until stuck.) Then the packing lemma ensures that S can't have too many vertices. (This is where the upper bound on the number of parts comes from.)

Now we partition the vertex set essentially by taking a Voronoi partition with respect to the distance between neighborhoods — we create parts $V_1, \ldots, V_t$ where for each vertex v, we place v into the part corresponding to the vertex s_i for which $|N(v) \mathbin{\Delta} N(s_i)|$ is smallest (this ensures $|N(v) \mathbin{\Delta} N(s_i)| \leq \delta$). (The proof actually does something slightly different — we go through all the vertices of S in order, and first take V_1 to be the set of all v for which $|N(v) \mathbin{\Delta} N(s_1)| \leq \delta$, then take V_2 to be the set of all remaining v with $|N(v) \mathbin{\Delta} N(s_2)| \leq \delta$, and so on; but the idea is the same.)



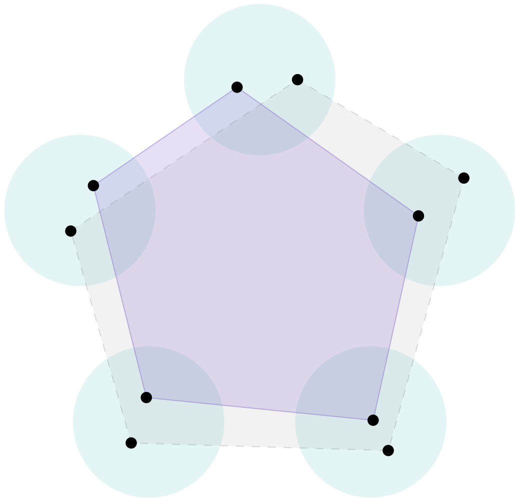
This won't actually give an equipartition, so we'll have to fix that — at some point we'll have to cut up these parts into small sets of fixed size. (We'll potentially have some small set of residual vertices left in each part when we do this; but we can do the classical trick of putting these things together and then splitting them up, and since there are very few such vertices we don't really have to care about them; they get absorbed into the ε-fraction of exceptional k-tuples.)

In order to show that this partition works, fix a k-tuple of parts $V_1, \ldots, V_k$, and assume for contradiction that its density is in $(\varepsilon, 1 - \varepsilon)$. The idea is to count pairs of k-tuples (e, e') where e and e' both consist of one vertex from each part and they differ in exactly one vertex, but e is an edge and e' isn't.

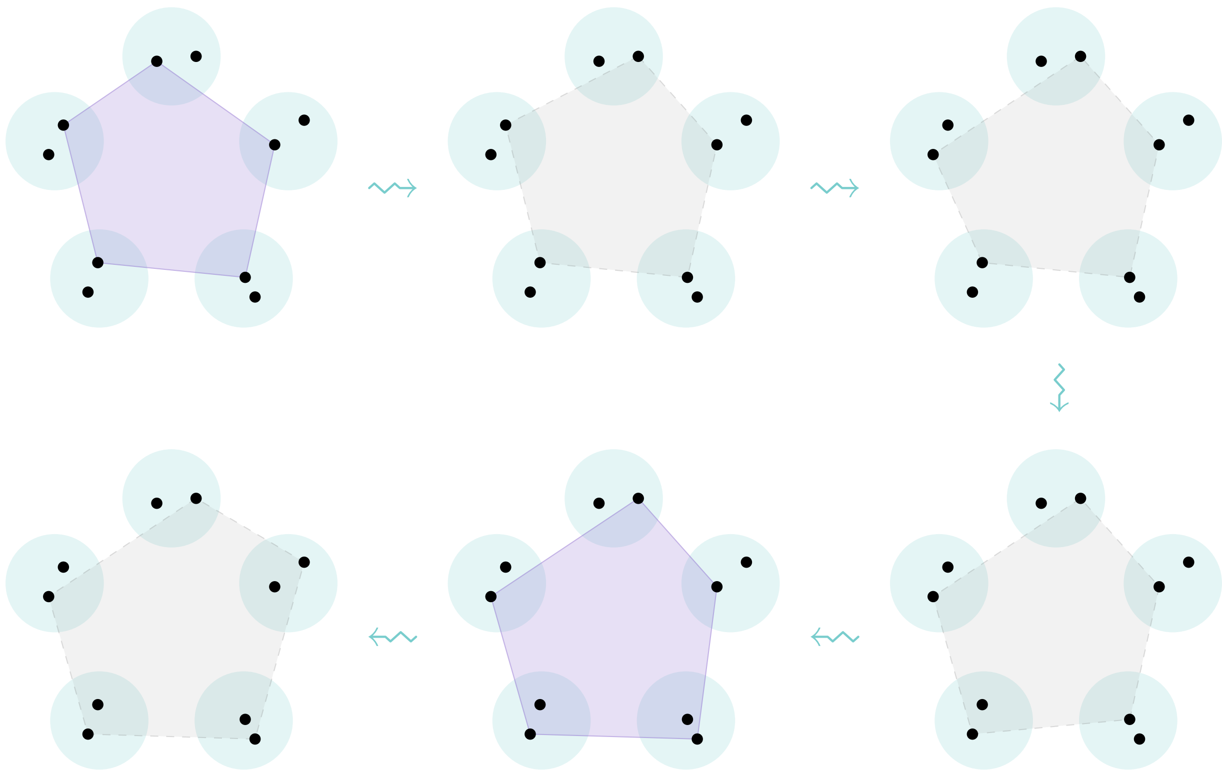


For example, if $k = 5$, then we have 5 parts, and we want to count the number of ways to choose one vertex from four of these parts and two vertices from the fifth, such that one of these 5-tuples forms an edge and the other doesn't (as shown above).

On one hand, if our density is in $(\varepsilon, 1-\varepsilon)$, then we claim that there are lots of such parts (at least around $\varepsilon^2 \left|V_i\right|^{k+1}$, where $|V_i|$ is the common size of each of the parts — in other words, if we choose e and e' arranged in this way at random (choosing one vertex from $k-1$ parts and two from the last part), then with reasonable probability one is an edge and the other isn't). To see this, imagine starting with *any* two k-tuples of vertices with one vertex in each part.



Then we can 'walk' from one of these k-tuples to the other by changing one vertex at a time. If the first k-tuple isn't an edge and the second is, then at some point during this process, we'll find e and e' which differ in exactly one vertex where one is an edge and the other isn't.



Then by using a union bound, we can get a lower bound on the probability that for a random pair (e, e') (sharing all but one vertex), one of e and e' is an edge and the other isn't.

On the other hand, we can also get an *upper* bound on the number of such (e, e') using the fact that the vertices in each part have similar neighborhoods — suppose we want to bound the number of pairs (e, e') where the one vertex in which they differ is in V_1. Imagine we *first* choose the vertices $u \in V_1$ and $v \in V_1$ for e and e', respectively. Then we need to count the number of ways to choose a $(k-1)$-tuple of one vertex from each of the remaining parts such that u forms an edge with this $(k-1)$-tuple, but v doesn't (or vice versa). But this means that $(k-1)$-tuple must be in $N(u) \,\Delta\, N(v)$, and since u and v are in the same part we have $|N(u) \,\Delta\, N(v)| \leq 2\delta$. So there are at most 2δ choices for the remaining $k-1$ vertices, which gives an upper bound on the number of such (e, e'). □

Remark 2.5. The only place where we use VC-dimension is to ensure that the number of parts in this partition is small. Also, here we use the full version of the packing lemma, where the constant doesn't depend on d.

§2.3 The Erdős–Hajnal conjecture

Now we'll see an application of this result; everything from now on is in the 2-uniform setting (where we have graphs instead of hypergraphs).

By Ramsey theory, we know that for every n-vertex graph G, there is an independent set or clique of size $\Omega(\log n)$ in G. The Erdős–Hajnal conjecture is about proving a stronger bound than this under a fairly general condition.

Conjecture 2.6 (Erdős–Hajnal) — Let $\mathcal{P}$ be any hereditary property which is not satisfied by all graphs. Then any graph with the property $\mathcal{P}$ has a clique or independent set of size at least $n^{\varepsilon(\mathcal{P})}$, where $\varepsilon(\mathcal{P})$ is some constant depending on the property $\mathcal{P}$.

This conjecture is still wide open. In the paper in which they made this conjecture, Erdős and Hajnal showed that this statement is true if we replace $n^{\varepsilon(\mathcal{P})}$ with $e^{\sqrt{\log n}}$. This bound remained the best-known bound for the general case (though there has been a lot of work for specific properties, and there do exist properties for which we know the conjecture is true — for example, graphs without a 5-cycle), but recently there have been small improvements — Bucić, Scott, Nguyen, and Seymour proved (in a series of 5 papers) that we can get the slightly stronger bound of $e^{\sqrt{\log n \log\log n}}$, among several other things.

Today we'll show that graphs with bounded VC-dimension *almost* satisfy this conjecture.

Theorem 2.7

A graph with bounded VC-dimension must have a clique or independent set of size $\exp(c(\log n)^{1-o(1)})$.

For comparison, in order to prove the Erdős–Hajnal conjecture we'd like to get rid of the $-o(1)$ term — $\exp(c(\log n))$ is the same thing as n^{ε}.

Proof sketch. We'll discuss the main ideas of the proof. The idea is to work with *cographs*.

Definition 2.8. A *cograph* is a graph that can be constructed by the following procedure: a single vertex is a cograph, and a graph formed by taking two cographs (not necessarily of the same size) and drawing either all or none of the edges between them is also a cograph.

A standard inductive argument shows that cographs are perfect graphs, meaning that $\omega(G) = \chi(G)$ — this means a n-vertex cograph has a clique or independent set of size at least $n^{1/2}$.

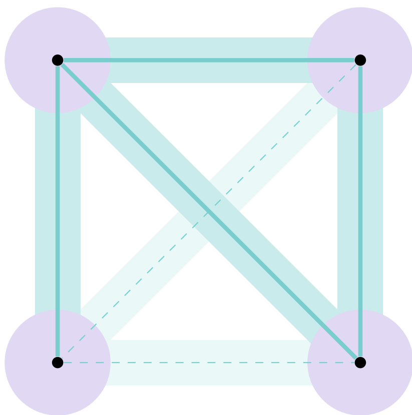
And at a very high level, the idea is that the regularity lemma will allow us to construct a large cograph inside our given graph G.

First, apply the regularity lemma; this gives us a bunch of parts such that nearly all pairs of parts have either very high or very low density. Then using this regularity partition, we can find a reasonably large set of vertices such that for all u and v in this set:

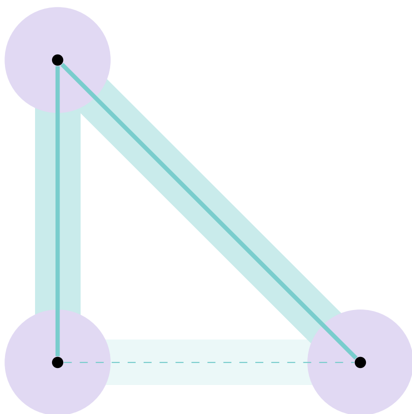
- If u and v are adjacent, then the density between their parts is very high.
- If u and v are *not* adjacent, then the density between their parts is very low.

One way to construct such a set is by using Turán's theorem — we can construct an auxiliary graph where we draw an edge between each pair (u, v) with this property. (We can imagine constructing this auxiliary graph by taking the original graph, removing all edges between irregular parts and within parts, removing all edges between low-density pairs, and adding in all edges between low-density pairs which were originally non-edges.) This graph has very high density, so by Turán's theorem we can find in it a large clique.

(It should also be possible to find such a set by taking a random set of representative points.)



Then inside this set of vertices, we can find a reasonably large cograph (using induction — taking subgraphs doesn't increase the VC-dimension).



Now if we take each vertex and replace it with a cograph inside its part (such that all the same edge relations hold — if we have an edge between the vertices in V_1 and V_2 in the original cograph, we should replace them with cographs in V_1 and V_2 such that all edges between these two cographs are present), then we'll get a bigger cograph.

In order to do this, we first take V_1 and clean it up in order to ensure that each of its vertices belongs to very few 'bad' pairs — i.e., edges between low-density pairs of parts, or non-edges between high-density pairs of parts. Since the total number of such bad pairs is very low, we can do this by pruning out at most half of the vertices in V_1 (in fact, we can do this by pruning out a very small fraction, but the bound of $\frac{1}{2}$ is good enough).

Once we've done this cleaning procedure to V_1, we then take a large cograph inside V_1 (which we can do by induction). And then we remove all vertices in the other parts which form bad pairs with any vertex in this cograph (e.g., if (V_1, V_2) has high density, then we remove all vertices in V_2 which are *not* adjacent to some vertex in this cograph, and if (V_1, V_2) has low density, then we remove all vertices in V_2 which *are* adjacent). This cograph is much smaller than the sizes of the parts, so this doesn't shrink those parts by too much. (The cograph is essentially of size $\exp(c(\log n)^{1-o(1)})$; we need the $o(1)$ for this to work.)

And then we can continue doing this (taking a cograph in the next part, and so on). □