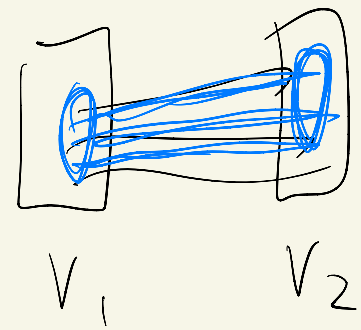
Hypergraph regularity and higher arity VC-dimension

joint with Henry Towsner

Graph: $G = (V_1, V_2, E)$, $E \subseteq V_1 \times V_2$

Induced subgraph: $H = (W_1, W_2, E\restriction_{W_1 \times W_2})$ for some $W_i \subseteq V_i$.



Intuitively for a fixed finite H,
every suff. random G contains
H as an induced subgraph.

Expectation some induced H is forbidden $\Rightarrow$ some structure on G.

Ex. $H = \bullet\!\!-\!\!\!-\!\!\bullet$, G is H-free $\Rightarrow G = (V_1, V_2, \emptyset)$.

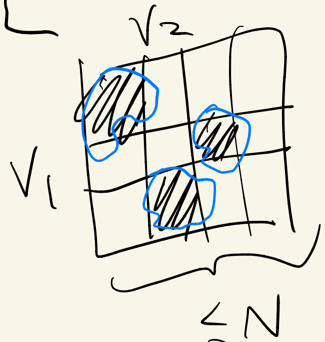
$H = \bullet \quad \bullet$, —//— $\Rightarrow G = (V_1, V_2, V_1 \times V_2)$

What about arbitrary H?

Fact [Alon, Fischer, Newman, '07], [Lovász – Szegedy '10]

Let H is any finite graph and $\varepsilon > 0$. Then $\exists\ N = N(H, \varepsilon)$ satisfying the following:

Let $G = (V_1, V_2, E)$ be any finite H-free graph. Then $\exists$ sets $A_1, \ldots, A_N \subseteq V_1$, $B_1, \ldots, B_N \subseteq V_2$ s.t. taking $E' = \bigsqcup_{1 \leq i \leq N} A_i \times B_i$, then $|E \triangle E'| \leq \varepsilon \cdot |V_1| \cdot |V_2|$.



Moreover: – $N = O\left(\left(\frac{1}{\varepsilon}\right)^d\right)$ for some $d = d(H) \in \mathbb{N}$.

– Each A_i, B_i in the Boolean algebra gen. by the fibers of E (for $a \in V_1$, $E_a = \{ b \in V_2 : (a, b) \in E \}$

— // —).

A k-hypergraph $G = (V_1, \ldots, V_k, E)$, $E \subseteq \prod_{i=1}^{k} V_i$

Induced k-hypergraphs: $H = (W_1, \ldots, W_k, E \restriction \prod_{i=1}^{k} W_i)$ for some $W_i \subseteq V_i$.

If $k < k'$, H a k-hypergraph, G a k'-hypergraph.

Then G omits H if for any fixed $(k'-k)$ coordinates, say $(a_{k+1}, \ldots, a_{k'}) \in V_{k+1} \times \cdots \times V_{k'}$, then the hypergraph of $\{(x_1, \ldots, x_k) \in \prod_{i=1}^{k} V_i : (x_1, \ldots, x_k, a_{k+1}, \ldots, a_{k'}) \in E\}$. omits H.

For $r < k$, $\boxed{B_{k,r}}$ is the Bool. algebra of subsets of $\prod_{i=1}^{k} V_i$ generated by "intersections of k-ary cylinder sets":

i.e. sets of the form

$$X = \{(x_i)_{1 \le i \le k} \in \prod_{i=1}^{k} V_i : \bigwedge_{S \subseteq [k], |S| \le r} (x_i)_{i \in S} \in X_S\},$$

$[k] = \{1, \ldots, k\}$

for some $X_S \subseteq \prod_{i \in S} V_i$.

$\iff$ the indicatator function χ_E is a product of function depending on at most r vars.

<u>Ex</u> $B_{2,1} = \sigma(A \times B)$, $A \subseteq V_1$, $B \subseteq V_2$.

<u>Thm</u> [C., Towsner] For any $k \leq k'$, finite k-hypergraph H and $\varepsilon > 0$, $\exists N = N(H, k, k', \varepsilon)$, satisfying: if G, a finite H-free k'-hypergraph, then for some $E' \subseteq \prod_{i=1}^{k'} V_i$ a union of at most N intersections of cylinder sets in $\boxed{B_{k',k}}$, $|E \Delta E'| \leq \varepsilon \cdot \prod_{i=1}^{k'} |V_i|$.

Note: $k' = 2$ — the previous ~~fact~~.

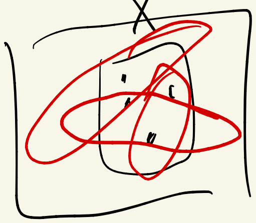
<u>Rem</u> — The case $k', k = 1$ — [C., Starchenko]

- For arbitrary measures on vertices, $[0,1]$-valued functions instead of graphs
- The cylinder sets are given by the fibers of E.

Connection to VC (Vapnik-Chervonenkis) theory

Def $\mathcal{F} \subseteq P(V)$ has VC-dim$(\mathcal{F}) = d$ if d is max s.t. $\exists X \subseteq V$, $|X| = d$ s.t. $\forall Y \subseteq X$, $\exists S \in \mathcal{F}$ s.t. $X \cap S = Y$.

V X



① [Sauer-Shelah] If $VC(\mathcal{F}) \leq d$, then $\forall n \in \mathbb{N}$, $\forall X \subseteq V$, $|X| = n$, then

$$\left| \{ Y \subseteq X : Y = X \cap S \text{ for some } S \in \mathcal{F} \} \right| = O(n^d).$$

② "Existence of ε-nets" [Haussler, Welzl]
$\forall d \in \mathbb{N}$, $\forall \varepsilon > 0$, $\exists N = N(\varepsilon, d)$ s.t. for any finite prob. space (V, μ) and $\mathcal{F} \subseteq P(V)$ with $VC(\mathcal{F}) \leq d$, $\exists x_1, \dots, x_N \in V$ s.t. $\forall S \in \mathcal{F}$, $\mu(S) > \varepsilon \Rightarrow x_i \in S$ for some i.

Given a graph $G = (V_1, V_2, E)$, $VC(G) = VC(\{E_y \subseteq V_1 : y \in V_2\})$.

Note: G omits $H = (W_1, W_2, E|_{W_1 \times W_2})$ with $|W_i| \leq d$ $\Rightarrow$ $VC(G) \leq 2d$

$VC(G) \leq d \Rightarrow G$ omits a certain H with parts of size d and 2^d

$\boxed{\text{Sketch for } k=2}$ Fix H, $\varepsilon > 0$.

Assume $G = (V_1, V_2, E)$, $VC(G) \le d$.

Then $VC\left(\underbrace{\{E_y \triangle E_{y'} \subseteq V_1 : y, y' \in V_2\}}\right) \le 10d$.

Let $x_1, \dots, x_n \in V_1$ be an ε-net for.

That is, $\forall y, y' \in V_2$, $\mu(E_y \triangle E_{y'}) > \varepsilon \Rightarrow x_i \in E_y \triangle E_{y'}$ for some i.

For each $S \subseteq \underbrace{\{x_1, \dots, x_n\}}$, let $B_S := \{y \in V_2 : x_i \in E_y \Leftrightarrow x_i \in S\}$.

$\Rightarrow$ for any $y, y' \in B_S$, $\mu(E_y \triangle E_y) \le \varepsilon$

— up to symm. diff. ε, there are only finitely many fibers of E.

Pick $b_S \in B_S$, let $E' := \bigcup \underbrace{E_{b_S}} \times \underbrace{B_S}$

and $\mu(E \triangle E') < \varepsilon$.

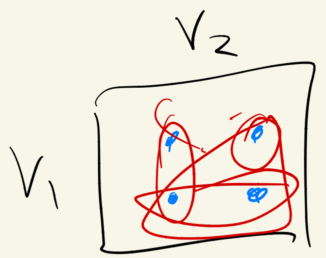
By Sauer-Shelah: only poly$\left(\frac{1}{\varepsilon}\right)$-many diff. B_S.

VC_k-dimension

$\mathcal{F} \subseteq P(V_1 \times \dots \times V_k)$

$VC_k(\mathcal{F})$ is the max d s.t. $\exists X_i \subseteq V_i$, $|X_i| = d$

s.t. $\forall Y \subseteq X_1 \times \dots \times X_k$, $\exists S \in \mathcal{F}$ s.t. $Y = \left(\prod_{i=1}^{k} X_i\right) \cap S$.



$k = 2$

$VC_1 = VC$

Ex $F, G, H \subseteq V^2$ are arbitrary

Define $E \subseteq V^3$: $(x,y,z) \in E \iff$ odd number of pairs $(x,y), (x,z), (y,z)$ belong to F, G, H resp.

Then $VC_2(E) < 100$.

[C., Palacin, Takeuchi '19] If $\mathcal{F} \subseteq P\left(\prod_{i=1}^{k} V_i\right)$, $VC_k(\mathcal{F}) \le d$, then $\exists \varepsilon = \varepsilon(d) > 0$ s.t. $\forall X_i \subseteq V_i$, $|X_i| = n$, there are at most $2^{n^{k-\varepsilon}}$ diff. sets in $\left\{\left(\prod_{i=1}^{k} X_i\right) \cap S : S \in \mathcal{F}\right\}$.

$\boxed{k-\varepsilon}$

(k', k).

Thm [C., Towsner] $\forall k, d, \varepsilon > 0$ $\exists \underline{N}$ satisfying:
let (V_i, μ_i) be finite prob. spaces, $i \in [k]$, and
$\mathcal{F} \subseteq \mathcal{P}(\prod_{i \in [k]} V_i)$ with $VC_k(\mathcal{F}) \leq d$. Then
$\exists S_1, \ldots, S_{\underline{N}} \in \mathcal{F}$ s.t. $\forall S \in \mathcal{F}$ we have
$\mu_1 \times \ldots \times \mu_k (S \triangle D) < \varepsilon$ for some D
given by a Bool. comb. of $S_1, \ldots, S_N$ and
N sets in $\mathcal{B}_{k, k-1}$ (depending on S).

$\uparrow$

In fact, $(\leq k-1)$-ary fibers of E.

<ins>Rem</ins> $k = 1$, $\mathcal{B}_{1,0} = \{\emptyset, V_1\}$.