# Removal Lemmas: Summer School 2025

## 5 Property testing

Let us consider the following equivalent form of the infinite removal lemma:

**Theorem 5.1** (Infinite removal lemma, sampling formulation)**.** *Let $\mathcal{H}$ be a family of graphs. For every $\varepsilon > 0$ there is an integer $q = q_{\mathcal{H}}(\varepsilon)$ such that if $G$ is $\varepsilon$-far from induced $\mathcal{H}$-free, then with probability at least 0.99, a sample of $q$ vertices from $G$ is not induced $\mathcal{H}$-free.*

To see that the above follows from Theorem 3.2, note that a sample of $m = m_{\mathcal{H}}(\varepsilon)$ vertices contains an induced copy of $H \in \mathcal{H}$ with probability at least $\delta = \delta_{\mathcal{H}}(\varepsilon)$, so a sample of $q = Cm/\delta$ contains such a copy with probability tending to 1 as $C$ tends to infinity. The reverse direction is also true, i.e., that Theorem 5.1 implies Theorem 3.2 (this is left as an exercise for the reader), and $q$ depends polynomially on $m, 1/\delta$.

Theorem 5.1 leads to the notion of *property testing*. A *property tester* for a graph property $\mathcal{P}$ is a randomized algorithm that distinguishes between graphs which satisfy $\mathcal{P}$ and graphs that are $\varepsilon$-far from $\mathcal{P}$, with success probability at least 0.99 (say) in both cases. Namely, if an input $G$ satisfies $\mathcal{P}$ then the algorithm must accept $G$ with probability at least 0.99, and if $G$ is $\varepsilon$-far from $\mathcal{P}$ then the algorithm must reject $\mathcal{P}$ with probability at least 0.99.[1] The algorithm works by sampling vertices and making edge queries, i.e., asking if a pair of vertices $u, v$ forms an edge. We require that the *sample complexity* of the algorithm, i.e., the number of vertices it samples, depends only on $\varepsilon$ and not on the size of the input graph $G$.

Property testing originated in the 1990s, and has since been thoroughly studied. The model we discuss here is called the *dense graph model*. There are also other models of property testing, e.g., for constant-degree graphs.

If $\mathcal{P}$ is *hereditary*, i.e., closed under the removal of vertices[2], then then there is a very simple tester for $\mathcal{P}$: Simply sample $q$ vertices of the input graph $G$, and accept if and only if the subgraph induced by the sample satisfies $\mathcal{P}$. As the property is hereditary, if $G$ satisfies $\mathcal{P}$ then the tester accepts with probability 1.[3] The fact that this algorithm is correct is simply the statement of Theorem 5.1.

One of the early and highly influential works on property testing is a paper of Goldreich, Goldwasser and Ron, where several natural graph properties were shown to be testable with polynomial sample complexity. Two key examples are $k$-colorability and having an independent set of size at least $\rho n$ (for a fixed $\rho \in [0, 1]$). Note that the latter is not a hereditary property. To illustrate some of the ideas in this work, let us show that bipartiteness is testable with sample complexity poly$(1/\varepsilon)$.

[1] In the intermediate range – i.e. that $G$ doesn't satisfy $\mathcal{P}$ but is $\varepsilon$-close to it – there is no requirement on the algorithm.

[2] Note that a graph property is hereditary if and only if it is defined in terms of a (possibly infinite) family of forbidden induced subgraphs.

[3] Accepting inputs satisfying $\mathcal{P}$ with probability 1 is known as having *one-sided error*.

**Theorem 5.2** (Goldreich-Goldwasser-Ron 1998)**.** *If $G$ is $\varepsilon$-far from being bipartite then a sample of $q = \tilde{O}(1/\varepsilon^2)$ vertices of $G$ induces a non-bipartite graph with probability at least* 0.9*.*

**Proof.** First, by deleting at most $\frac{\varepsilon}{2}n^2$ edges, we may pass to a (spanning) subgraph of $G$ where every vertex has degree 0 or at least $\frac{\varepsilon}{2}n$. Indeed, as long as there is a vertex $v$ with degree at least 1 but less than $\frac{\varepsilon}{2}n$, delete all edges touching $v$. Note that the remaining graph (after the edge deletions) is $\frac{\varepsilon}{2}$-far from bipartiteness (because $G$ is $\varepsilon$-far from bipartiteness). Let $U$ be the set of vertices which are not isolated.

The key idea is to sample in two stages. First, sample vertices $x_1, \dots, x_s$, $s = \frac{2}{\varepsilon}\log(\frac{100}{\varepsilon})$. For a vertex $u \in U$, the probability that $u$ has no neighbors in $X = \{x_1, \dots, x_s\}$ is at most

$$\left(1 - \frac{\varepsilon}{2}\right)^s \leq e^{-\varepsilon s/2} \leq \frac{\varepsilon}{100}.$$

Let $U'$ be the set of $u \in U$ which have a neighbor in $X$. By the above and Markov's inequality, with probability at least 0.95 we have $|U'| \geq |U| - \frac{\varepsilon}{5}n$. Make the vertices in $U \setminus U'$ isolated by deleting at most additional $\frac{\varepsilon}{5}n^2$ edges.

If $G[X]$ is not bipartite then we are already done. Otherwise, fix any bipartition $(A_1, A_2)$ of $G[X]$. For $i = 1, 2$, let $U_i$ be the set of all vertices $u \in U'$ which have a neighbor in $A_i$ (if $u$ has a neighbor in both $A_1, A_2$ then place $u$ in one of the sets $U_1, U_2$ arbitrarily). Then $U' = U_1 \cup U_2$. Also, every vertex of $G$ which is not currently isolated belongs to $U_1 \cup U_2$. As the remaining graph is $\frac{\varepsilon}{4}$-far from bipartiteness, there are at least $\frac{\varepsilon}{4}n^2$ edges which are inside $U_1$ or inside $U_2$. Now sample additional vertices $Y = \{y_1, \dots, y_t\}$, $t = \frac{100s}{\varepsilon} = \tilde{O}(1/\varepsilon^2)$. Note that if we sample any of the edges inside $U_1$ or $U_2$, then the bipartition $(A_1, A_2)$ of $G[X]$ cannot be extended to a bipartition of $G[X \cup Y]$. The probability that we sample no such edge is at most

$$\left(1 - \frac{\varepsilon}{4}\right)^{t/2} \leq e^{-\varepsilon t/8} < 0.05 \cdot 2^{-s}.$$

By taking a union bound over all at most $2^s$ bipartitions $(A_1, A_2)$ of $G[X]$, we see that the probability that $G[X \cup Y]$ is bipartite is at most 0.1, as required. ■

We note that the bound on $q$ in Theorem 5.2 has been improved to $\tilde{O}(1/\varepsilon)$, which is optimal up to the logarithmic terms. Also, such a bound has been proved for much more general testing tasks, such as testing hypergraph $k$-colorability and (more generally) testing satisfiability.

We now move on to testing for independent sets; more precisely, for the property of containing an independent set of size at least $\rho n$. As mentioned above, this property was shown to be testable already by Goldreich, Goldwasser and Ron. However, very recently, a new proof was discovered by Blais and Seth, which uses the container method and supplies optimal bounds on the sample complexity of such a tester. Here we present a version of their argument with somewhat weaker bounds, for the sake of simplicity.

**Theorem 5.3** (Blais-Seth 2023)**.** *Let $G$ be an $n$-vertex graph which is $\varepsilon$-far from containing an independent set of size at least $\rho n$. Then with probability at least 0.9, a sample $X = \{x_1, \dots, x_q\}$ of $q = \tilde{O}(1/\varepsilon^3)$ vertices from $G$ satisfies that $\alpha(G[X]) \leq (\rho - \frac{\varepsilon}{4})q$.*

**Proof.** The assumption on $G$ implies that every vertex-set $U \subseteq V(G)$ of size at least $\rho n$ contains at least $\varepsilon n^2$ edges. But in fact we can get a bit more: every vertex set $U$ of size at least $(\rho - \frac{\varepsilon}{2})n$ contains at least $\frac{\varepsilon}{2}n^2$ edges. Indeed, otherwise, add arbitrary $\frac{\varepsilon}{2}n$ vertices to $U$, delete all (at most

$\frac{\varepsilon}{2}n^2$ edges) touching these vertices, and delete all edges inside $U$. This gives an independent set of size at least $\rho n$, in contradiction to our assumption.

The key part of the argument is the following claim, which uses the container algorithm:

**Claim 5.4.** *Set $t := \frac{1}{\varepsilon}$. For every independent set $I \subseteq V(G)$ with $|I| > t$, there are sets $F = F(I) \subseteq I$ and $C = C(I) \subseteq V(G)$ such that $I \subseteq F \cup C$, $|F| \leq t$ and $|C| \leq (\rho - \frac{\varepsilon}{2})n$. Furthermore, $C$ depends only on $F$ (and not on $I$). More precisely, if $F = F(I)$, then $C(I) = C(F)$.*[4]

**Proof.** The container algorithm is as follows: Initialize $F_0 = \emptyset$ and $C_0 = V(G)$. For $i \geq 0$, if $|C_i| \leq (\rho - \frac{\varepsilon}{2})n$ or $C_i \cap I = \emptyset$, then stop and output $F := F_i, C := C_i$. Otherwise proceed as follows:

1. Order the elements of $C_i$ as $v_1, \dots, v_m$, such that for every $1 \leq j \leq m$, $v_j$ has maximum degree in $G[\{v_j, \dots, v_m\}]$.[5]

2. Let $1 \leq j \leq m$ be minimal with $v_j \in I$.

3. Move $v_j$ to $F$, and delete from $C_i$ the vertices $v_1, \dots, v_j$ and all neighbors of $v_j$ in $\{v_{j+1}, \dots, v_m\}$. Namely, set
$$F_{i+1} = F_i \cup \{v_j\}$$
$$C_{i+1} = C_i \setminus \{v_k : k \leq j \text{ or } v_j v_k \in E\}.$$

It is easy to see that $F_i \subseteq I$ and $I \subseteq F_i \cup C_i$ throughout the process, and the upper bound $|C| \leq (\rho - \frac{\varepsilon}{2})n$ is guaranteed by the process. The fact that $C$ depends only on $F$ is a standard fact about the container algorithm, and is left for the reader. It now suffices to show that the process stops in at most $t$ steps, as this would guarantee the upper bound on $F$. Suppose otherwise. Consider any step $i$ in the process except the last. Since the process did not stop at step $i+1$, we have $|C_{i+1}| \geq (\rho - \frac{\varepsilon}{2})n$. Considering the ordering $v_1, \dots, v_m$ at step $i$ (see Item 1 above), we have $C_{i+1} \subseteq \{v_{j+1}, \dots, v_m\}$, so in particular, the set $U := \{v_j, \dots, v_m\}$ has size at least $(\rho - \frac{\varepsilon}{2})n$. As explained above, this means that $U$ contains at least $\frac{\varepsilon}{2}n^2$ edges. As $v_j$ is chosen as a vertex of maximum degree in $G[U]$, we have $d_U(v_j) \geq \varepsilon n$. Hence, at least $\varepsilon n$ vertices are removed from $C_i$ at this step. As this is true for every step except the last, and we assumed that the process lasts at least $t$ steps, we get $|C_{t-1}| \leq n - (t-1) \cdot \varepsilon n < (\rho - \frac{\varepsilon}{2})n$, a contradiction to the assumption that the process did not stop before step $t$. ■

Let us now prove Theorem 5.3. Sample vertices $X = \{x_1, \dots, x_q\}$ uniformly at random and independently.[6] We need to upper-bound the probability that $G[X]$ contains an independent set of size at least $(\rho - \frac{\varepsilon}{4})q$. Fix any such set $I$, and let $F = F(I)$ and $C = C(F)$ be given by Claim 5.4. If $I \subseteq X$, then $F \subseteq X$ and $X$ contains at least $|I| - |F|$ vertices of $C$. Note that $|I| - |F| \geq (\rho - \frac{\varepsilon}{4})q - \frac{1}{\varepsilon} \geq (\rho - \frac{\varepsilon}{3})q$ (provided that $\varepsilon$ is small enough). Thus, we see that if $\alpha(G[X]) \geq (\rho - \frac{\varepsilon}{4})q$, then there is a set $F \subseteq X = \{x_1, \dots, x_q\}$ such that $|X \cap C(F)| \geq (\rho - \frac{\varepsilon}{3})q$. We union bound over the choice of indices in $[q]$ which play the role of $F$, and then condition on the outcome of these indices (i.e., we condition on $F$). The number of choices for the index set is $\binom{q}{\leq t} \leq e^{O(\frac{1}{\varepsilon}\log q)}$. Having conditioned on $F$ and setting $C = C(F)$, the random variable $|X \cap C|$ is

[4] The set $F$ is usually called the *fingerprint* of $I$, and the set $C$ is called the *container* corresponding to $F$.

[5] If two vertices have the same degree, then ties are broken according to a pre-fixed ordering on $V(G)$.

[6] Here we sample with repetition for the sake of simplicity. Alternatively, one can sample a subset $X \subseteq V(G)$ of size $q$ uniformly at random and use concentration inequalities for the hypergeometric distribution.

distributed as $\mathrm{Bin}(q, |C|/n)$, and so has expectation at most $(\rho - \frac{\varepsilon}{2})q$. By Hoeffding's inequality, the probability that $|X \cap C| \geq (\rho - \frac{\varepsilon}{3})q$ is at most $e^{-\Omega(q\varepsilon^2)}$. Combining the above, we see that

$$\mathbb{P}\left[\alpha(G[X]) \geq \left(\rho - \frac{\varepsilon}{4}\right) q\right] \leq e^{O(\frac{1}{\varepsilon}\log q)} \cdot e^{-\Omega(q\varepsilon^2)},$$

which is less than 0.1 if $q = \frac{C}{\varepsilon^3} \log \frac{1}{\varepsilon}$ for a large enough constant $C$. ■

Using Theorem 5.3, we can now obtain a tester for large independent sets.

**Theorem 5.5** (Golreich-Goldwasser-Ron 1998, Blais-Seth 2023 (optimal bound)). *The property of containing an independent set of size at least $\rho n$ is testable with sample complexity poly$(1/\varepsilon)$.*

**Proof.** The algorithm samples vertices $X = \{x_1, \ldots, x_q\}$, where $q = \tilde{O}(1/\varepsilon^3)$ is given by Theorem 5.3, and accepts if and only if $\alpha(G[X]) > (\rho - \frac{\varepsilon}{4})q$. If $G$ has an independent set $I$ of size at least $\rho n$, then one can show using Hoeffding's inequality that $|X \cap I| > (\rho - \frac{\varepsilon}{4})q$ with probability at least 0.9. And if $G$ is $\varepsilon$-far from containing an independent set of size at least $\rho n$, then $\alpha(G[X]) \leq (\rho - \frac{\varepsilon}{4})q$ with probability at least 0.9 by Theorem 5.3. ■

# 6 Hypergraph regularity and VC-dimension for hypergraphs

In this section we consider the extensions of the notions of regularity and VC-dimension to hypergraphs. For simplicity, we consider 3-uniform hypergraphs, but all material covered in this section extends to higher uniformity.

## 6.1 Regularity

What is the appropriate notion of regularity for 3-uniform hypergraphs? A natural attempt is as follows: A 3-partite 3-graph $\mathcal{H} = (X, Y, Z)$ is *$\varepsilon$-regular* if for every $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$ with $\frac{|X'|}{|X|}, \frac{|Y'|}{|Y|}, \frac{|Z'|}{|Z|} \geq \varepsilon$, it holds that $|d(X', Y', Z') - d(X, Y, Z)| \leq \varepsilon$, where

$$d(X, Y, Z) := \frac{e(X, Y, Z)}{|X||Y||Z|}$$

is the density of $(X, Y, Z)$. This notion of regularity is called *weak regularity* (for reasons that we shall see shortly). One can indeed prove a regularity lemma with respect to this notion; the statement and its proof are straightforward generalizations of Szemerédi's regularity lemma. The problem, however, is that this notion of regularity is not strong enough to imply a counting lemma. To see this, consider the following key example: Take random bipartite graphs $E \subseteq X \times Y$, $F \subseteq X \times Z$ and $G \subseteq Y \times Z$ (with edge-probability $\frac{1}{2}$, say, though this will not be important). Define the hypergraph $\mathcal{H}$ to consist of all triangles formed by $E, F, G$, i.e., $xyz \in E(\mathcal{H})$ if $xy \in E, xz \in F, yz \in G$. This hypergraph $\mathcal{H}$ is weakly regular. Indeed, for all linear-sized sets $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$, the density of triangles between $(X', Y', Z')$ is very close to that of $(X, Y, Z)$, because the graphs $E, F, G$ are random. However, $\mathcal{H}$ does not contain a tri-induced copy of every fixed-size 3-partite 3-graph.[7] Indeed, let $\mathcal{K}$ be the 3-partite 3-graph obtained from $K^{(3)}_{2,2,2}$ by deleting one edge. If, by contradiction,

[7] A tri-induced copy of a 3-partite 3-graph $\mathcal{K} = (A, B, C)$ is defined in the natural way: it is an injection $\varphi : V(\mathcal{K}) \to V(\mathcal{H})$ such that $\varphi(A) \subseteq X, \varphi(B) \subseteq Y, \varphi(C) \subseteq Z$, and for every $(a, b, c) \in A \times B \times C$, $abc \in E(\mathcal{K})$ if and only if $\varphi(a)\varphi(b)\varphi(c) \in E(\mathcal{H})$.

$x_1, x_2, y_1, y_2, z_1, z_2$ form a tri-induced copy of $\mathcal{K}$ in $\mathcal{H}$ (with $x_1, x_2 \in X$ and so on), then all pairs $x_i y_j$ $(1 \le i, j \le 2)$ belong to $E$, and similarly all pairs $x_i z_j$ belong to $F$ and all pairs $y_i z_j$ to $G$. But then this is a copy of $K^{(3)}_{2,2,2}$ instead of an induced copy of $\mathcal{K}$. So we see that the analogue of Lemma 1.7 fails for weak regularity.[8]

Let us now introduce a stronger notion of regularity which does admit a counting lemma. Consider a tripartite graph with parts $X, Y, Z$ consisting of $E \subseteq X \times Y$, $F \subseteq X \times Z$ and $G \subseteq Y \times Z$. Let $\triangle(E, F, G)$ denote the set of triangles in $G$, i.e., the set of triples $(x, y, z) \in X \times Y \times Z$ with $xy \in E, xz \in F, yz \in G$. Now let $\mathcal{H}$ be a 3-partite 3-graph on $X, Y, Z$. The *density* of $\mathcal{H}$ with respect to $(E, F, G)$ is defined as

$$d(\mathcal{H} \mid E, F, G) := \frac{|E(\mathcal{H}) \cap \triangle(E, F, G)|}{|\triangle(E, F, G)|}.$$

Namely, the density is the fraction of $(E, F, G)$-triangles which are edges of $\mathcal{H}$. The definition of regularity in the 3-graph regularity lemma is with respect to this density. That is, we require that for every $E' \subseteq E, F' \subseteq F, G' \subseteq G$, if $\triangle(E', F', G') \ge \varepsilon \triangle(E, F, G)$, it holds that

$$\left| d(\mathcal{H} \mid E', F', G') - d(\mathcal{H} \mid E, F, G) \right| \le \varepsilon.$$

In fact, the known proof of the 3-graph counting lemma requires a somewhat stronger version of the above: We say that $\mathcal{H}$ is $(\varepsilon, r)$*-regular with respect to* $(E, F, G)$ if for every $E_1, \dots, E_r \subseteq E, F_1, \dots, F_r \subseteq F, G_1, \dots, G_r \in G$ with $\sum_{i=1}^{r} \triangle(E_i, F_i, G_i) \ge \varepsilon \triangle(E, F, G)$, it holds that

$$\left| \frac{|E(\mathcal{H}) \cap \bigcup_{i=1}^{r} \triangle(E_i, F_i, G_i)|}{|\bigcup_{i=1}^{r} \triangle(E_i, F_i, G_i)|} - d(\mathcal{H} \mid E, F, G) \right| \le \varepsilon.$$

In order to make use of the $\varepsilon$-regularity of $\mathcal{H}$ with respect to $(E, F, G)$, we have to be able to count the $(E, F, G)$-triangles. To this end, we require that the bipartite graphs $E, F, G$ themselves are regular.[9]

The 3-graph regularity lemma supplies vertex partitions[10] of $X, Y, Z$ as well as pair partitions of $X' \times Y', X' \times Z', Y' \times Z'$ for any choice of vertex-parts $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$. The lemma guarantees that for "most"[11] choices of vertex-parts $X', Y', Z'$ and pair-parts $E \subseteq X' \times Y', F \subseteq X' \times Z', G \subseteq Y' \times Z'$, it holds that $E, F, G$ are $\delta$-regular (for a suitable small enough $\delta$) and $\mathcal{H}$ is $(\varepsilon, r)$-regular with respect to $E, F, G$.[12]

Just as in the graph case, the proof of the 3-graph regularity lemma proceeds via density increment: if $\mathcal{H}$ is not $\varepsilon$-regular with respect to many triples of pair-parts $(E, F, G)$, then one can refine the pair partition and thus increase the energy function. One then needs to apply the graph regularity lemma to the new pair parts to maintain the property that all pair parts are regular. This in turn refines the vertex partition. The repeated applications of graph regularity result in a *Wowzer-type* bound. The wowzer function is the iterated tower function, i.e., $\text{wowzer}(x) = \text{tower}(\text{wowzer}(x - 1))$.

[8] Another classical example to the counting lemma failing is as follows. Take a random tournament $T$ on $n$ vertices, and consider the 3-uniform hypergraph $\mathcal{H}$ on $V(T)$ whose edges are the cyclic triangles in $T$. It can be shown that $\mathcal{H}$ is weakly-regular (due to the fact that $T$ is random), but any 4 vertices of $\mathcal{H}$ contain at most 2 edges (because any 4 vertices in a tournament contain at most 2 cyclic triangles). In particular, $\mathcal{H}$ does not contain $K_4^{(3)}$ (or even $K_4^{(3)} - e$).

[9] I.e., they should be regular enough so that we may apply the graph counting lemma (Lemma 1.3). This means that the degree of regularity should be small enough as a function of the densities of $E, F, G$.

[10] The goal of the vertex partitions is to make the parts of the pair partitions regular.

[11] "Most" means the following: If we sample $(x, y, z) \in X \times Y \times Z$ uniformly at random and consider the unique vertex- and pair-parts containing $(x, y, z)$, then these have the desired property with probability at least $1 - \varepsilon$. In other words, $(E, F, G)$ is weighted by $\frac{|\triangle(E,F,G)|}{|X||Y||Z|}$.

[12] For the proof of the counting lemma, the parameter $r$ must also depend on (i.e., be large enough with respect to) the densities of $E, F, G$.

## 6.2 VC-dimension

Recall that a shattered set in a graph is a vertex-set $X = \{x_1, \dots, x_d\}$ such that for every $I \subseteq [d]$, there is a vertex $y_I$ such that $\{i \in [d] : x_i y_I \in E\} = I$. In 3-uniform hypergraphs, a shattered set will consist of pairs instead of vertices; the rest of the definition is very similar. Namely, a set of pairs $\{e_1, \dots, e_d\}$ in a 3-graph $\mathcal{H}$ (so $e_i \in \binom{V(\mathcal{H})}{2}$ for every $1 \le i \le d$) is *shattered* if for every $I \subseteq [d]$, there is a vertex $y_I \in V(\mathcal{H})$ such that $e_i \cup \{y_I\} \in E(\mathcal{H})$ if and only if $i \in I$.

As $e_1, \dots, e_d$ are now pairs (instead of vertices), they themselves carry structure, i.e., of a graph. Hence, we can have different definitions of VC-dimension depending on the structure of shattered sets which we are considering.

**Definition 6.1.**

1. *The* strong VC-dimension *of $\mathcal{H}$ is the maximum size of a shattered set of pairs $e_1, \dots, e_d$ (here there are no restrictions on $e_1, \dots, e_d$).*[13]

2. *The* $\mathrm{VC}_1$-dimension *(also known as* slicewise VC-dimension*) of $\mathcal{H}$ is the maximum size of a shattered set $e_1, \dots, e_d$ which forms a star.*

3. *The* $\mathrm{VC}_2$-dimension *of $\mathcal{H}$ is the maximum size of a shattered set $e_1, \dots, e_d$ which forms a complete bipartite graph.*

For convenience, in what follows we often consider 3-partite 3-graphs (instead of general 3-graphs), but all material applies to general 3-graphs as well.

Fox, Pach and Suk proved that hypergraphs with bounded strong VC-dimension have small homogeneous partitions:

**Theorem 6.2** (Fox-Pach-Suk 2019)**.** *If $\mathcal{H} = (X, Y, Z)$ has strong VC-dimension $d$, then it has an $\varepsilon$-homogeneous equipartition of size at most $(1/\varepsilon)^{O(d)}$. Namely, there are equipartitions $X = X_1 \cup \dots \cup X_t$, $Y = Y_1 \cup \dots \cup Y_t$, $Z = Z_1 \cup \dots \cup Z_t$, where $t \le (1/\varepsilon)^{O(d)}$, such that for all but at most $\varepsilon t^3$ of the triples $(i, j, k) \in [t]^3$ it holds that $d(X_i, Y_j, Z_k) \le \varepsilon$ or $d(X_i, Y_j, Z_k) \ge 1 - \varepsilon$.*

Thus, strong VC-dimension behaves similarly to the graph case.

Let us now consider $\mathrm{VC}_1$- and $\mathrm{VC}_2$-dimension. Note that if a hypergraph $\mathcal{H}$ has unbounded $\mathrm{VC}_2$-dimension, then it contains a tri-induced copy of every 3-partite 3-graph $\mathcal{K} = (A, B, C)$. Indeed, first map $A \times B$ onto a shattered complete bipartite graph, denoting the mapping by $\varphi$, and then, for every $c \in C$, take a vertex $z_c \in V(\mathcal{H})$ which makes an edge precisely with the pairs $\{\varphi(a)\varphi(b) : abc \in E(\mathcal{K})\}$. Thus, in this sense, $\mathrm{VC}_2$-dimension is analogous to the graph case: bounded $\mathrm{VC}_2$-dimension is equivalent to excluding tri-induced copies of a fixed 3-partite 3-graph, just as bounded (graph) VC-dimension is equivalent to excluding bi-induced copies of a fixed bipartite graph. By the same considerations, bounded $\mathrm{VC}_1$-dimension is equivalent to excluding tri-induced copies of a fixed 3-partite 3-graph $\mathcal{K} = (A, B, C)$ where $|A| = 1$. In other words, bounded $\mathrm{VC}_1$-dimension is equivalent to all link graphs having bounded VC-dimension.[14]

It turns out that bounded $\mathrm{VC}_1$-dimension (which is a weaker assumption than bounded strong VC-dimension) also implies the existence of $\varepsilon$-homogeneous vertex-partitions. This was first proved by Chernikov and Towsner, without any quantitative bound on the size of the partition. A double exponential bound $2^{2^{\mathrm{poly}(1/\varepsilon)}}$ was subsequently proved by Terry. Very recently, this was improved to an exponential bound:

[13] The term "strong VC-dimension" is not standard. I have not found a better name for this definition.

[14] The link $L_{\mathcal{H}}(x)$ of a vertex $x$ is the graph $\{yz : xyz \in E(\mathcal{H})\}$.

**Theorem 6.3** (Gishboliner-Shapira-Wigderson)**.** *If $\mathcal{H} = (X, Y, Z)$ has bounded $VC_1$-dimension, then it has an $\varepsilon$-homogeneous equipartition of size at most $2^{poly(1/\varepsilon)}$.*

A construction by Terry shows that an exponential bound is best possible.[15] The following remains open:

**Conjecture 6.4.** *If $\mathcal{H} = (X, Y, Z)$ has bounded $VC_1$-dimension, then there are $X' \subseteq X, Y' \subseteq Y, Z' \subseteq Z$ with $\frac{|X'|}{|X|}, \frac{|Y'|}{|Y|}, \frac{|Z'|}{|Z|} \geq poly(\varepsilon)$ such that $d(X', Y', Z') \leq \varepsilon$ or $d(X', Y', Z') \geq 1 - \varepsilon$.*

**Proof sketch of Theorem 6.3.** For simplicity, suppose that $|X| = |Y| = |Z| = n$. Fix any pair $(x, y) \in X \times Y$. Consider the link $L_{\mathcal{H}}(x)$, which is a bipartite graph between $Y$ and $Z$. As $\mathcal{H}$ has bounded VC$_1$-dimension, $L_{\mathcal{H}}(x)$ has bounded VC-dimension. Hence, by (the bipartite version of) Lemma 4.4, there is a partition $Y = Y_1^{(x)} \cup \cdots \cup Y_s^{(x)}$, $s = \text{poly}(1/\varepsilon)$, such that two vertices $y_1, y_2$ in the same part satisfy $|N_Z(y_1) \triangle N_Z(y_2)| \leq \varepsilon n$, where the neighborhoods are in $L_{\mathcal{H}}(x)$. In other words, $|N_Z(x, y_1) \triangle N_Z(x, y_2)| \leq \varepsilon n$, where the neighborhoods are in $\mathcal{H}$. For simplicity, suppose that the partition $Y_1^{(x)} \cup \cdots \cup Y_s^{(x)}$ is an equipartition (this can be easily arranged by allowing one exceptional part). Also, without loss of generality, suppose that $y \in Y_1^{(x)}$. Pick any $y' \in Y_1^{(x)}$, and now consider $L_{\mathcal{H}}(y)$, which is a bipartite graph between $X$ and $Z$. By the same argument as above, we get an equipartition $X = X_1^{(y')} \cup X_s^{(y')}$ such that two vertices $x_1, x_2$ in the same part satisfy $|N_Z(x_1, y') \triangle N_Z(x_2, y')| \leq \varepsilon n$. Without loss of generality, $x \in X_1^{(y')}$. Now, for every $x' \in X_1^{(y')}$, we have, by the triangle inequality:

$$|N_Z(x', y') \triangle N_Z(x, y)| \leq |N_Z(x', y') \triangle N_Z(x, y')| + |N_Z(x, y') \triangle N_Z(x, y)| \leq 2\varepsilon n.$$

Also, the number of choices for $(x', y')$ is $(n/s)^2$. Summarizing, for every pair[16] $(x, y) \in X \times Y$, there are at least $(n/s)^2$ pairs $(x', y') \in X \times Y$ with $|N_Z(x', y') \triangle N_Z(x, y)| \leq 2\varepsilon n$.

Now sample $f_1, \dots, f_r \in X \times Y$ uniformly at random, where $r = s^2 \log \frac{1}{\varepsilon}$, and define $E_i := \{(x, y) \in X \times Y : |N_Z(x, y) \triangle N_Z(f_i)| \leq 2\varepsilon n\}$. For every $i \in [r]$, very two pairs in $F_i$ have the same neighborhood in $Z$, up to an error of $4\varepsilon n$ (by the triangle inequality). Also, for a given $(x, y) \in X \times Y$, we have

$$\mathbb{P}[(x, y) \notin E_1 \cup \cdots \cup E_r] \leq \left(1 - \frac{1}{s^2}\right)^r \leq \varepsilon.$$

In conclusion, we get a partition $X \times Y = E_0 \cup E_1 \cup \cdots \cup E_r$ such that $|E_0| \leq \varepsilon n^2$, and for every $i \in [r]$ and $(x, y) \in E_i$, $|N_Z(x, y) \triangle N_Z(f_i)| \leq 2\varepsilon n$.

Now, let $Z_i := N_Z(f_i)$, and let $\mathcal{P}_Z$ be the Venn diagram of the sets $Z_1, \dots, Z_r$. This is a partition of $Z$ into at most $2^r = 2^{\text{poly}(1/\varepsilon)}$ sets.[17] We expect a homogeneous behavior between each $E_i$ $(1 \leq i \leq r)$ and a typical part of $\mathcal{P}_Z$.

The proof proceeds by repeating the above argument for $X \times Z$ and $Y \times Z$. In the former case we obtain a partition of $X \times Z$ and a partition $\mathcal{P}_Y$ of $Y$, and in the latter case we obtain a partition of $Y \times Z$ and a partition $\mathcal{P}_X$ of $X$. One can now show that $(\mathcal{P}_X, \mathcal{P}_Y, \mathcal{P}_Z)$ is an $\varepsilon$-homogeneous partition of $\mathcal{H}$.[18] We omit the details. ■

[15] This construction is as follows: Partition $X, Y$ into equal-sized parts $X_1, \dots, X_k$ and $Y_1, \dots, Y_k$, respectively, where $k = (1/\varepsilon)^{0.1}$, say. For each $i = 1, \dots, k$, take a uniformly random subset $Z_i \subseteq Z$ and add all edges in $X_i \times Y_i \times Z_i$ (the sets $Z_1, \dots, Z_k$ are chosen independently). One can show that any $\varepsilon$-homogeneous partition of this hypergraph has size at least $2^{(1/\varepsilon)^{\Omega(1)}}$.

[16] In fact, we can only guarantee this for almost every pair, because of the aforementioned exceptional sets. But we ignore this technicality.

[17] This step is where the exponential bound in Theorem 6.3 comes from.

[18] More precisely, $\varepsilon'$-homogeneous for some $\varepsilon' = \varepsilon^c$, $c > 0$ constant.

Moving to $\mathrm{VC}_2$-dimension, what can we say about a hypergraph with bounded $\mathrm{VC}_2$-dimension? Recall the construction described in Section 6.1 (arising from random graphs). This construction has bounded $\mathrm{VC}_2$-dimension but no $\varepsilon$-homogeneous partition (even for $\varepsilon = 0.49$) of size independent of $n$. So bounded $\mathrm{VC}_2$-dimension does not imply the existence of (bounded-size) $\varepsilon$-homogeneous vertex partitions. Observe, however, that this construction does have a homogeneous *pair partition*: Partition $X \times Y$ into $E_1 := E, E_2 := X \times Y \setminus E$, and similarly partition $X \times Z$ into $F_1, F_2$ and $Y \times Z$ into $G_1, G_2$. Then for every $i, j, k = 1, 2$, the hypergraph $\mathcal{H}$ is homogeneous (i.e., complete or empty) over $\triangle(E_i, F_j, G_k)$. It turns out that this is a general phenomenon:

**Theorem 6.5** (Chernikov-Towsner 2020)**.** *If $\mathcal{H} = (X, Y, Z)$ has bounded $VC_2$-dimension, then there are equipartitions $X \times Y = E_1 \cup \cdots \cup E_t$, $X \times Z = F_1 \cup \cdots \cup F_t$, $Y \times Z = G_1 \cup \cdots \cup G_t$, where $t$ depends only on $\varepsilon$, such that for all but $\varepsilon t^3$ of the triples $(E_i, F_j, G_k)$ it holds that $d(\mathcal{H} \mid E_i, F_j, G_k) \leq \varepsilon$ or $d(\mathcal{H} \mid E_i, F_j, G_k) \geq 1 - \varepsilon$.*

Theorem 6.5 can be deduced from the hypergraph regularity lemma, as follows: Taking a regular partition, one can show that for a regular triple $(E_i, F_j, G_k)$, the density of $\mathcal{H}$ over $\triangle(E_i, F_j, G_k)$ is at most $\varepsilon$ or at least $1 - \varepsilon$. Indeed, otherwise $\mathcal{H}$ contains a tri-induced copy of every fixed 3-partite 3-graph, by a counting lemma analogous to Lemma 1.7.

In fact, the result of Chernikov and Towsner is more general. For each uniformity $k \geq 2$ and $1 \leq \ell \leq k - 1$, they define a suitable notion of $\mathrm{VC}_\ell$-dimension, and show that a $k$-graph with bounded $\mathrm{VC}_\ell$-dimension has an *$\varepsilon$-homogeneous partition of uniformity $\ell$*, i.e., a partition of all $\ell$-sets.

**Problem 6.6.** *Does Theorem 6.5 hold with $t = poly(1/\varepsilon)$?*