Property testing for bipartite patterns

Yufei Zhao (MIT)

Joint work with Noga Alon (Princeton) and Jacob Fox (Stanford)

June, 2018

Testing triangle-free-ness

[Rubinfield and Sudan '96] [Goldreich, Goldwasser, Ron '98]

Goal: determine if an n-vertex graph is triangle-free or ϵ-far from triangle free

ϵ-far from triangle-free: need to delete $\geq \epsilon n^2$ edges to make it triangle-free.

Testing triangle-free-ness

[Rubinfield and Sudan '96] [Goldreich, Goldwasser, Ron '98]

Goal: determine if an n-vertex graph is triangle-free or ϵ-far from triangle free

ϵ-far from triangle-free: need to delete $\geq \epsilon n^2$ edges to make it triangle-free.

Algorithm: Sample $C(\epsilon)$ triples at random

- If never see a triangle, then output "triangle-free"
- Else, output "ϵ-far from triangle-free"

Testing triangle-free-ness

[Rubinfield and Sudan '96] [Goldreich, Goldwasser, Ron '98]

Goal: determine if an n-vertex graph is triangle-free or ϵ-far from triangle free

ϵ-far from triangle-free: need to delete $\geq \epsilon n^2$ edges to make it triangle-free.

Algorithm: Sample $C(\epsilon)$ triples at random

- If never see a triangle, then output "triangle-free"
- Else, output "ϵ-far from triangle-free"

Theorem (Ruzsa–Szemerédi 1976)

For every $\epsilon > 0$ there exists a $C(\epsilon) > 0$ so that this algorithm succeeds with probability $> 2/3$ (one-sided error)

Testing triangle-free-ness

[Rubinfield and Sudan '96] [Goldreich, Goldwasser, Ron '98]

Goal: determine if an n-vertex graph is triangle-free or ϵ-far from triangle free

ϵ-far from triangle-free: need to delete $\geq \epsilon n^2$ edges to make it triangle-free.

Algorithm: Sample $C(\epsilon)$ triples at random

- If never see a triangle, then output "triangle-free"
- Else, output "ϵ-far from triangle-free"

Theorem (Ruzsa–Szemerédi 1976)

For every $\epsilon > 0$ there exists a $C(\epsilon) > 0$ so that this algorithm succeeds with probability $> 2/3$ (one-sided error)

Proof: By Szemerédi's graph regularity lemma. Gives $C(\epsilon) = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ height poly$(1/\epsilon)$

Remark: False with $C(\epsilon) = \text{poly}(1/\epsilon)$

Testing sum-free-ness

Goal: determine if $A \subset G$ (abelian group) is sum-free or or ϵ-far from sum-free

sum-free: no solutions to $x + y = z$

ϵ-far from sum-free: need to remove $\geq \epsilon|G|$ elements to make sum-free

Testing sum-free-ness

Goal: determine if $A \subset G$ (abelian group) is sum-free or or ϵ-far from sum-free

sum-free: no solutions to $x + y = z$

ϵ-far from sum-free: need to remove $\geq \epsilon|G|$ elements to make sum-free

Algorithm: Sample $C(\epsilon)$ triples $(x, y, x + y) \in G^3$ at random

- If never see $x, y, x + y \in A$, then output "sum-free"
- Else, output "ϵ-far from sum-free"

Testing sum-free-ness

Goal: determine if $A \subset G$ (abelian group) is sum-free or or ϵ-far from sum-free

sum-free: no solutions to $x + y = z$
ϵ-far from sum-free: need to remove $\geq \epsilon|G|$ elements to make sum-free

Algorithm: Sample $C(\epsilon)$ triples $(x, y, x + y) \in G^3$ at random

- If never see $x, y, x + y \in A$, then output "sum-free"
- Else, output "ϵ-far from sum-free"

Theorem (Green 2005)

For every $\epsilon > 0$ there exists a $C(\epsilon) > 0$ so that this algorithm succeeds with probability $> 2/3$ (one-sided error)

Testing sum-free-ness

Goal: determine if $A \subset G$ (abelian group) is sum-free or or ϵ-far from sum-free

sum-free: no solutions to $x + y = z$
ϵ-far from sum-free: need to remove $\geq \epsilon|G|$ elements to make sum-free

Algorithm: Sample $C(\epsilon)$ triples $(x, y, x + y) \in G^3$ at random

- If never see $x, y, x + y \in A$, then output "sum-free"
- Else, output "ϵ-far from sum-free"

Theorem (Green 2005)

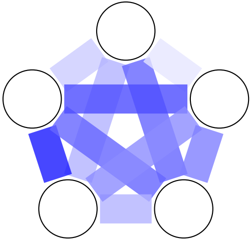
For every $\epsilon > 0$ there exists a $C(\epsilon) > 0$ so that this algorithm succeeds with probability $> 2/3$ (one-sided error)

Proof: By regularity lemma. Gives $C(\epsilon) = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ height poly$(1/\epsilon)$
Remark: $C(\epsilon) = \text{poly}(1/\epsilon)$ works if $G = \mathbb{F}_p^n$ with p fixed [Fox–Lovász '17], but not for $G = \mathbb{Z}/N\mathbb{Z}$

Spoiler

For testing bipartite patterns, poly$(1/\epsilon)$ samples suffice

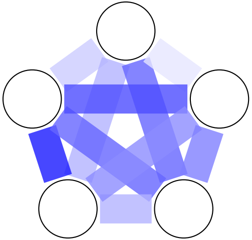
Szemerédi's graph regularity lemma



Graph regularity lemma

For every $\epsilon > 0$ there exists $M = M(\epsilon)$ so that every graph has a vertex partition into $\leq M$ parts so that all but $< \epsilon$ fraction of pairs are ϵ-regular

Szemerédi's graph regularity lemma



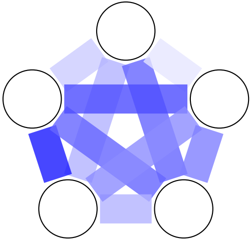
Graph regularity lemma

For every $\epsilon > 0$ there exists $M = M(\epsilon)$ so that every graph has a vertex partition into $\leq M$ parts so that all but $< \epsilon$ fraction of pairs are ϵ-regular

Graph removal lemma

For every $\forall\epsilon > 0$ and graph H there is some $\delta = \delta(H, \epsilon) > 0$ so that every n-vertex graph with H-density $< \delta$ can be made H-free by removing $< \epsilon n^2$ edges

Szemerédi's graph regularity lemma



Graph regularity lemma

For every $\epsilon > 0$ there exists $M = M(\epsilon)$ so that every graph has a vertex partition into $\leq M$ parts so that all but $< \epsilon$ fraction of pairs are ϵ-regular

Graph removal lemma

For every $\forall \epsilon > 0$ and graph H there is some $\delta = \delta(H, \epsilon) > 0$ so that every n-vertex graph with H-density $< \delta$ can be made H-free by removing $< \epsilon n^2$ edges

- $M(\epsilon) = 2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$ tower of height $\epsilon^{-O(1)}$ (cannot be improved [Gowers])
- Removal lemma holds with $\delta = M^{-O(1)} = 1/2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$
(possibly could be improved, but not beyond $\epsilon^{C \log(1/\epsilon)}$ when $H = K_3$)

When can you guarantee poly($1/\epsilon$) bounds?

When can you guarantee poly($1/\epsilon$) bounds?

For a graph with bounded VC dimension:

- Vertices can be partitioned into $\epsilon^{-O(1)}$ parts
- All but ϵ-fraction of pairs of vertex parts have densities $\leq \epsilon$ or $\geq 1 - \epsilon$

[Alon–Fischer–Newman, Lovász–Szegedy]

What is VC dimension?

Let $\mathcal{S}$ be a collection of subsets of Ω

$$\dim_{\mathrm{VC}}\mathcal{S} := \text{size of the largest shattered subset of } \Omega$$

$U \subset \Omega$ is shattered if for every $U' \subseteq U$ there exists $T \in \mathcal{S}$ such that $T \cap U = U'$

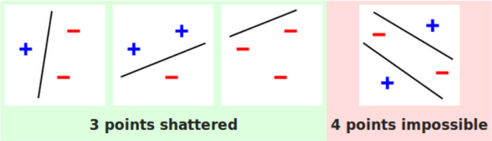
What is VC dimension?

Let $\mathcal{S}$ be a collection of subsets of Ω

$$\dim_{\text{VC}} \mathcal{S} := \text{size of the largest shattered subset of } \Omega$$

$U \subset \Omega$ is shattered if for every $U' \subseteq U$ there exists $T \in \mathcal{S}$ such that $T \cap U = U'$

E.g., the VC-dimension of the collection of half-planes in $\mathbb{R}^2$ is 3



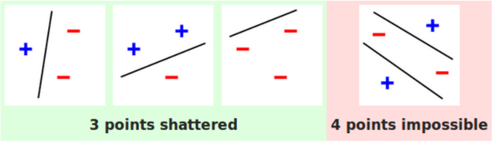
What is VC dimension?

Let $\mathcal{S}$ be a collection of subsets of Ω

$$\dim_{\mathrm{VC}} \mathcal{S} := \text{size of the largest shattered subset of } \Omega$$

$U \subset \Omega$ is shattered if for every $U' \subseteq U$ there exists $T \in \mathcal{S}$ such that $T \cap U = U'$

E.g., the VC-dimension of the collection of half-planes in $\mathbb{R}^2$ is 3



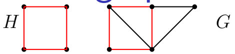
VC dimension of a graph G is defined to be the VC dimension of the collection of vertex neighborhoods ($\Omega = V(G)$):

$$\dim_{\mathrm{VC}} G := \dim_{\mathrm{VC}} \{N(v) : v \in V(G)\}$$

Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

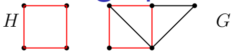
Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

H as a subgraph of G (all edges of H are present in G)

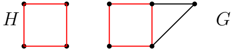


Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

H as a subgraph of G (all edges of H are present in G)

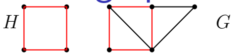


H as an induced subgraph of G (all edges of H are present in G and non-edges are not present)

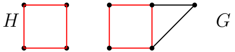


Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

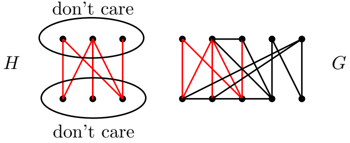
H as a subgraph of G (all edges of H are present in G)



H as an induced subgraph of G (all edges of H are present in G and non-edges are not present)

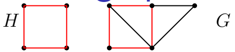


Bipartite H as a bi-induced subgraph (similar to induced but don't care about edges inside each bipartition)

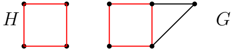


Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

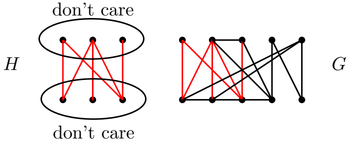
H as a subgraph of G (all edges of H are present in G)



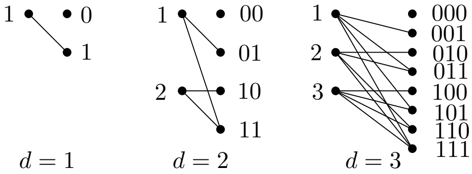
H as an induced subgraph of G (all edges of H are present in G and non-edges are not present)



Bipartite H as a bi-induced subgraph (similar to induced but don't care about edges inside each bipartition)

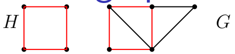


$\dim_{VC} G < d \iff G$ forbids the following as a bi-induced subgraph:

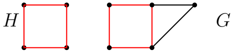


Bounded VC dimension $\iff$ forbidding a bi-induced subgraph

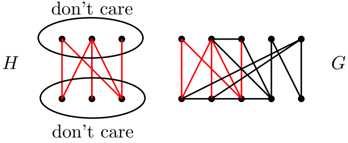
H as a subgraph of G (all edges of H are present in G)



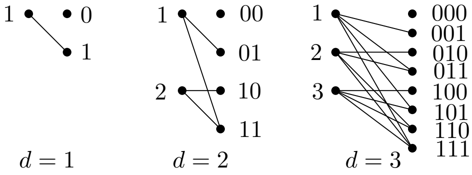
H as an induced subgraph of G (all edges of H are present in G and non-edges are not present)



Bipartite H as a bi-induced subgraph (similar to induced but don't care about edges inside each bipartition)



$\dim_{\mathrm{VC}} G < d \iff G$ forbids the following as a bi-induced subgraph:



Conversely, if G is bi-induced-H-free, then $\dim_{\mathrm{VC}} G = O_H(1)$

When can you guarantee poly($1/\epsilon$) bounds?

Regularity lemma for graphs of bounded VC dimension

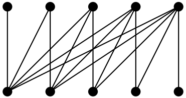
For a fixed bipartite H, if G is bi-induced-H-free, then G has a vertex partition into $\epsilon^{-O(1)}$ parts so that all but $\leq \epsilon$-fraction of pairs have edge-densities $\leq \epsilon$ or $\geq 1 - \epsilon$.

When can you guarantee poly$(1/\epsilon)$ bounds?

Regularity lemma for graphs of bounded VC dimension

For a fixed bipartite H, if G is bi-induced-H-free, then G has a vertex partition into $\epsilon^{-O(1)}$ parts so that all but $\leq \epsilon$-fraction of pairs have edge-densities $\leq \epsilon$ or $\geq 1 - \epsilon$.

A graph is *k-stable* if it does not contain a bi-induced half-graph on $2k$ vertices.

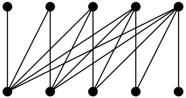


When can you guarantee poly($1/\epsilon$) bounds?

Regularity lemma for graphs of bounded VC dimension

For a fixed bipartite H, if G is bi-induced-H-free, then G has a vertex partition into $\epsilon^{-O(1)}$ parts so that all but $\le \epsilon$-fraction of pairs have edge-densities $\le \epsilon$ or $\ge 1-\epsilon$.

A graph is *k-stable* if it does not contain a bi-induced half-graph on $2k$ vertices.



Stable regularity lemma [Malliaris–Shelah]

If the graph is k-stable, then we can furthermore guarantee that **every** pair of parts has density $\le \epsilon$ or $\ge 1-\epsilon$.

Arithmetic setting

G abelian group, $A \subset G$

$$\dim_{\mathsf{VC}} A := \dim_{\mathsf{VC}} \{A + x : x \in G\} = \dim_{\mathsf{VC}} \mathsf{CayleyGraph}(G, A)$$

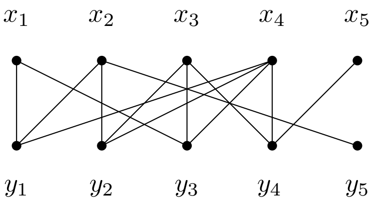
Arithmetic setting

G abelian group, $A \subset G$

$$\dim_{\mathrm{VC}} A := \dim_{\mathrm{VC}}\{A + x : x \in G\} = \dim_{\mathrm{VC}} \mathrm{CayleyGraph}(G, A)$$

We say that A contains a bi-induced copy of a bipartite graph $H = (U \cup V, E)$ if the same is true for CayleyGraph(G, A), or equivalently, there exists $x_1, \ldots, x_{|U|}, y_1, \ldots, y_{|V|} \in G$ such that

$$\forall (i,j) \in U \times V : \qquad x_i + y_j \in A \quad \text{if and only if } (i,j) \in E$$



Regularity lemmas with constraints

Graph regularity:

- Bounded VC-dimension (i.e., forbidding a bi-induced subgraph) [Alon–Fischer–Newman]: vertex-partition into $\leq \epsilon^{-O(1)}$ parts so that all but $\leq \epsilon$-fraction of pairs have edge-densities $\leq \epsilon$ or $\geq 1 - \epsilon$
- Stability (i.e., forbidding a fixed-size half-graph) [Malliaris–Shelah]: furthermore **every** pair of parts has density $\leq \epsilon$ or $\geq 1 - \epsilon$

Regularity lemmas with constraints

Graph regularity:

- Bounded VC-dimension (i.e., forbidding a bi-induced subgraph) [Alon–Fischer–Newman]: vertex-partition into $\leq \epsilon^{-O(1)}$ parts so that all but $\leq \epsilon$-fraction of pairs have edge-densities $\leq \epsilon$ or $\geq 1-\epsilon$
- Stability (i.e., forbidding a fixed-size half-graph) [Malliaris–Shelah]: furthermore **every** pair of parts has density $\leq \epsilon$ or $\geq 1-\epsilon$

Arithmetic regularity:

$A \subset G = \mathbb{F}_p^n$, p fixed (general groups: Conant–Pillay–Terry, Terry–Wolf)

- Stability [Terry–Wolf]: there exists a subspace $H \leq G$ with $[G:H] \leq e^{\epsilon^{-O(1)}}$ such that for all $x \in G$,

$$|A \cap (H+x)| \leq \epsilon |H| \text{ or } \geq (1-\epsilon)|H| \qquad (\dagger)$$

Regularity lemmas with constraints

Graph regularity:

- Bounded VC-dimension (i.e., forbidding a bi-induced subgraph) [Alon–Fischer–Newman]: vertex-partition into $\leq \epsilon^{-O(1)}$ parts so that all but $\leq \epsilon$-fraction of pairs have edge-densities $\leq \epsilon$ or $\geq 1-\epsilon$
- Stability (i.e., forbidding a fixed-size half-graph) [Malliaris–Shelah]: furthermore **every** pair of parts has density $\leq \epsilon$ or $\geq 1-\epsilon$

Arithmetic regularity:

$A \subset G = \mathbb{F}_p^n$, p fixed (general groups: Conant–Pillay–Terry, Terry–Wolf)

- Stability [Terry–Wolf]: there exists a subspace $H \leq G$ with $[G:H] \leq e^{\epsilon^{-O(1)}}$ such that for all $x \in G$,

$$|A \cap (H+x)| \leq \epsilon |H| \text{ or } \geq (1-\epsilon)|H| \qquad (\dagger)$$

- Bounded VC-dimension [Alon–Fox–Z.]: there exists a subspace $H \leq G$ with $[G:H] \leq \epsilon^{-O(1)}$ such that $(\dagger)$ holds for all but an $\leq \epsilon$-fraction of $x \in G$

Applications to removal lemma and property testing

Recall the graph removal lemma: $\forall \epsilon \exists \delta$: if an n-vertex graph has $< \delta n^3$ triangles, and it can be made triangle-free by removing $< \epsilon n^2$ edges.

Applications to removal lemma and property testing

Recall the graph removal lemma: $\forall \epsilon \exists \delta$: if an n-vertex graph has $< \delta n^3$ triangles, and it can be made triangle-free by removing $< \epsilon n^2$ edges.

Arithmetic removal lemma for bi-induced patterns [Alon–Fox–Z.]

Fix p and bipartite graph F. For every $\epsilon > 0$, there exists $\delta = \epsilon^{O(|V(F)|^3)}$ such that if the bi-induced-F-density in $A \subset \mathbb{F}_p^n$ is $< \delta$, then A can be made bi-induced-F-free by adding/deleting $< \epsilon p^n$ elements.

Corollary: property testing for arithmetic bi-induced patterns

Using poly$(1/\epsilon)$ samples, one can distinguish, with probability $> 2/3$, subsets that are bi-induced-F-free from those that are ϵ-far from bi-induced-F-free.

More generally holds for abelian groups with bounded exponent

Open questions

- Bounds for Conant–Pillay–Terry regularity lemmas for general groups with stability/bounded VC dimension hypotheses?
  (Proved via model theory. No bounds known. Maybe $\epsilon^{-O(1)}$?)
  - Quantitative bounds for stability in abelian groups recently proved by Terry–Wolf

Open questions

- Bounds for Conant–Pillay–Terry regularity lemmas for general groups with stability/bounded VC dimension hypotheses?
  (Proved via model theory. No bounds known. Maybe $\epsilon^{-O(1)}$?)
  - Quantitative bounds for stability in abelian groups recently proved by Terry–Wolf
- Property testing for bi-induced arithmetic patterns in a general (abelian) group?
  (No theorem known. Maybe $\epsilon^{-O(1)}$ bounds?)

Open questions

- Bounds for Conant–Pillay–Terry regularity lemmas for general groups with stability/bounded VC dimension hypotheses?
  (Proved via model theory. No bounds known. Maybe $\epsilon^{-O(1)}$?)
  - Quantitative bounds for stability in abelian groups recently proved by Terry–Wolf
- Property testing for bi-induced arithmetic patterns in a general (abelian) group?
  (No theorem known. Maybe $\epsilon^{-O(1)}$ bounds?)
- Property testing for induced arithmetic patterns?
  (No general theorem known, even in $\mathbb{F}_p^n$)