Regularity method for sparse graphs and its applications

Yufei Zhao (MIT)

Joint work with

David Conlon (Caltech)

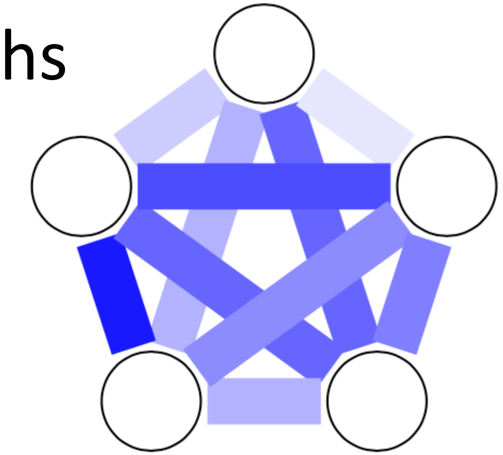
Jacob Fox (Stanford)

Benny Sudakov (ETH Zurich)

arXiv:2004.10180

Szemerédi's regularity method

A powerful rough structural description of all large graphs



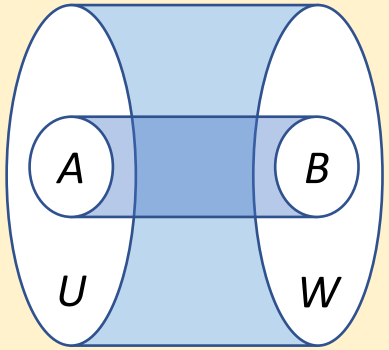
Graph regularity lemma (Szemerédi '70s).
For every $\varepsilon > 0$, every graph has an ε-regular vertex partition into $\leq M(\varepsilon)$ parts.

Many important applications:

- Extremal graph theory
- Additive combinatorics

(U,W) is ε-regular if
$\forall A \subset U, |A| \geq \varepsilon|U|$
$\forall B \subset W, |B| \geq \varepsilon|W|$
$$|d(A,B) - d(U,W)| \leq \varepsilon$$

An equitable vertex partition is ε-regular if at all but $\leq \varepsilon$-fraction of pairs of vertex sets are ε-regular



Graph removal lemmas

Graphs with "few" triangles can be made triangle-free by deleting "few" edges

Triangle removal lemma (Ruzsa–Szemerédi '76)**.** Every n-vertex graph with $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges

Graph removal lemma. Fix a graph H. Every n-vertex graph with $o(n^{v(H)})$ triangles can be made H-free by removing $o(n^2)$ edges

Applications:

- Extremal combinatorics/graph theory
- Property testing
- Additive combinatorics

Preview: sparse removal lemmas

n vertices, edge density on the order of $p = p(n) = o(1)$

"Sparse graph removal lemma". Fix a graph H. [Additional hypotheses]
An n-vertex graph with with $o(p^{e(H)} n^{v(H)})$ copies of H can be made H-free by removing $o(pn^2)$ edges.

Restricting to C_4-free graphs: edge-density $\lesssim p := 1/\sqrt{n}$ by Kővári–Sós–Turán

A new sparse $\{C_3, C_5\}$-removal lemma (Conlon, Fox, Sudakov, Z.)**.**
Every n-vertex C_4-free graph with $o(p^5 n^5) = o(n^{5/2})$ C_5's can be made C_5-free and C_3-free by removing $o(pn^2) = o(n^{3/2})$ edges.

Avoiding equations

i.e., avoiding $x + y = 2z$

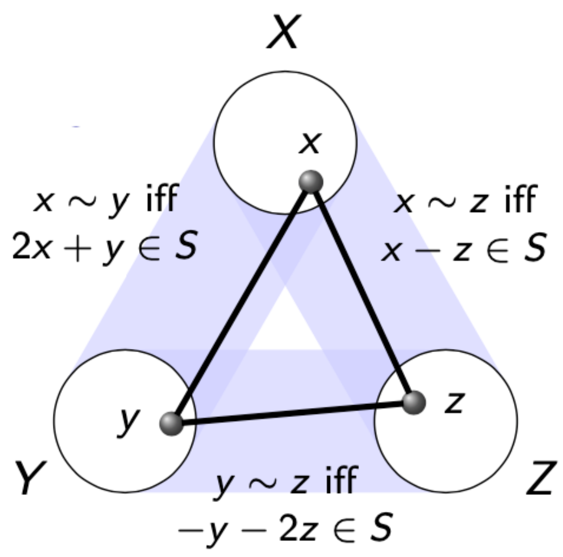
Roth's theorem ('53). Every subset of $[N] = \{1, 2, \ldots, N\}$ without a 3-term arithmetic progression (3-AP) has size $o(N)$.

Proof (Ruzsa–Szemerédi '76).

Given a 3-AP-free set S, set up a Cayley-like graph where every edge lies in exactly one triangle.

Apply triangle removal lemma on this graph to deduce that it has $o(N^2)$ edges.

Since $\#\text{edges} \asymp N|S|$, conclude $|S| = o(N)$

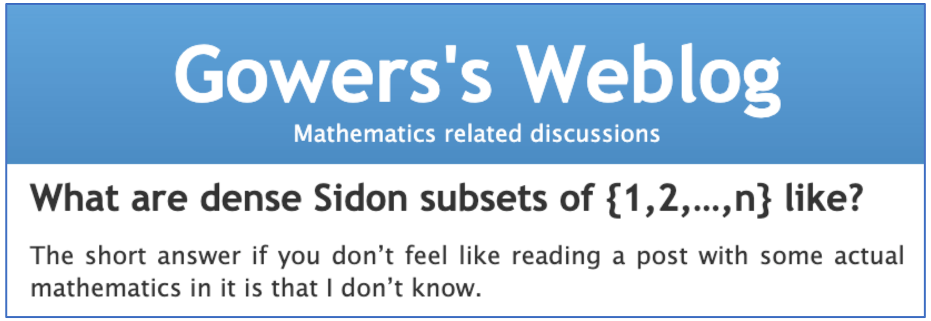


Preview: equation-avoidance in Sidon sets

A Sidon set is a set of integers avoiding nontrivial solutions to

$$x + y = z + w$$

Max size of a Sidon subset of $[N]$ is $\sim\sqrt{N}$



Question: What can we say about Sidon sets of nearly maximum size?

Ideally a description similar to Freiman's theorem, but seems a bit hopeless

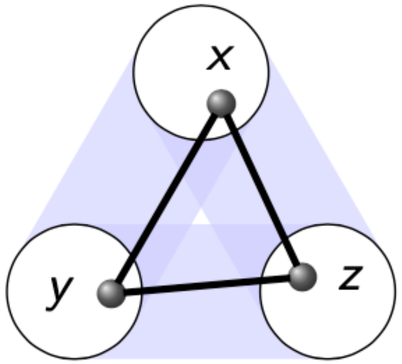
We give a weak answer: a Sidon set of size $\geq c\sqrt{N}$ contains solutions to every 5-variable translation-invariant linear equations with integer coefficients.

Theorem (Conlon, Fox, Sudakov, Z.)**.** Every Sidon subset of $[N]$ avoiding nontrivial solutions to $x_1 + x_2 + x_3 + x_4 = 4x_5$ has size $o\left(\sqrt{N}\right)$.

Sparse regularity

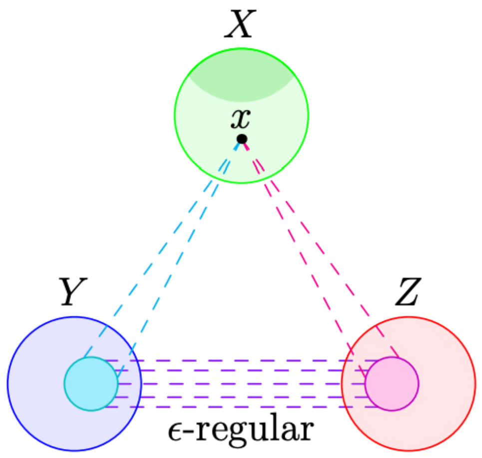
- Szemerédi's regularity lemma, in its original form, is useless for sparse graphs, i.e., with edge-density $o(1)$
- Sparse regularity: error tolerance commensurate with edge-density
- Obtaining a regularity partition for sparse graphs
  - Kohayakawa, Rödl ('90s): under additional hypothesis of "no dense spots"
  - Scott ('11): no additional hypothesis on the graph, but possibly hiding most of the graph in irregular pairs
- Applications? Counting lemma?

What is a counting lemma?



Given three “regular pairs” from the regularity partition, we want:

triangle density ≈ product of edge densities



We call such a statement a triangle counting lemma

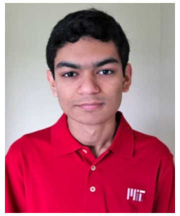
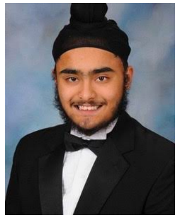
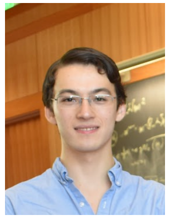
True for dense graphs

Serious challenges for sparser graphs (false without addl. hyp.)

Failures of counting lemmas in sparse graphs

Examples of random-like graphs without random-like triangle-counts

- $G(n, p)$ minus all triangles when $p = o(1/\sqrt{n})$ so that $p^3n^3 = o(pn^2)$
- Alon's pseudorandom triangle-free graph

- (with Ashwin Sah, Mehtaab Sawhney, and Jonathan Tidor arXiv: 2003.05272)

  Recent counterexample to Bollobás–Riordan conjectures on sparse graph limits, showing a strong failure of Chung—Graham—Wilson for sparse graph sequences:

  n vertices, edge density $n^{-o(1)}$, and normalized H-density → exp(– #△'s of H)

  (C_4-pseudorandom $\not\Rightarrow$ C_3-pseudorandom)

Sparse regularity applications

Green–Tao theorem ('08)**.** The primes contain arbitrarily long APs.

"Relative Szemerédi theorem." Fix k. Suppose $S \subset \mathbb{Z}/N\mathbb{Z}$ satisfies some *pseudorandomness hypotheses*. Then every k-AP-free subset of S has size $o(|S|)$.

Significantly simplified in [Conlon, Fox, Z. '15] via a new counting lemma

Additional hypothesis in this sparse counting lemma:

G is contained in some *pseudorandom host*

Removal lemmas

Triangle removal lemma (Ruzsa–Szemerédi ’76)**.** Every n-vertex graph with $o(n^3)$ triangles can be made triangle-free by removing $o(n^2)$ edges

$p := 1/\sqrt{n}$

A new sparse $\{C_3, C_5\}$-removal lemma (Conlon, Fox, Sudakov, Z.)**.** Every n-vertex graph with

- ~~no C_4~~ $o(p^4n^4) = o(n^2)$ C_4’s
- & $o(p^5n^5) = o(n^{5/2})$ C_5’s

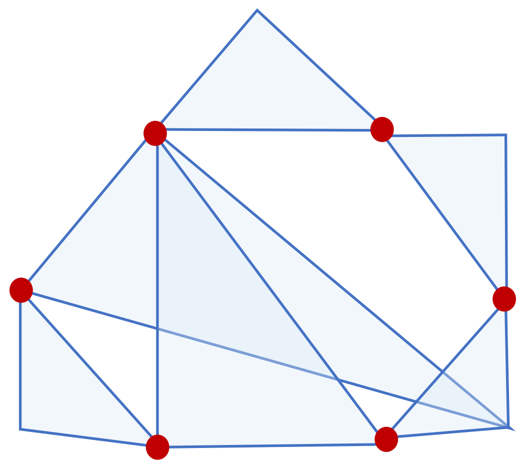
can be made C_5-free and C_3-free by removing $o(pn^2) = o(n^{3/2})$ edges.

Corollary. An n-vertex ~~C_5-free~~ graph can be made triangle-free by deleting $o(n^{3/2})$ edges.

with $o(n^2)$ C_5’s

$o(n^2)$ cannot be replaced by $o(n^{2.442})$
but we don’t know the optimal exponent

Extremal results in hypergraphs



In a hypergraph a Berge cycle of length k consists of

- k distinct vertices $v_1, \ldots, v_k$
- k distinct edges $e_1, \ldots, e_k$
- $v_i, v_{i+1} \in e_i\ \forall\ i$ (indices mod k)

Question. Max # edges in n-vertex 3-graph with no Berge cycle of length ≤ 5?

Previously: $O(n^{3/2})$ [Lazebnik, Verstraëte '03] [Ergemlidze, Methuku '18+]

Corollary of new result: $o(n^{3/2})$ (also same answer for r-graphs for all $r \geq 3$)

Also: # n-vertex 3-graphs with no Berge cycle of length ≤ 5 is $2^{o(n^{3/2})}$

Brown—Erdős—Sós type problems

BES(n, e, v) = max # triples in an n-vertex 3-graph without e edges spanning $\leq v$ vertices?

Ruzsa—Szemerédi theorem: BES(n, 6, 3) = $o(n^2)$

BES conjecture: BES(n,7,4) = $o(n^2)$,   BES(n, 8,5) = o(n^2), …

Corollary of new result: BES(10, 5) = $o(n^{3/2})$

Avoiding solutions to equations

Roth's theorem ('53). Every subset of $[N] = \{1, 2, \ldots, N\}$ without a 3-AP has size $o(N)$.

Theorem (CFSZ). Every subset of $[N]$ without a nontrivial solution to

$$x_1 + x_2 + 2x_3 = x_4 + 3x_5$$

has size $o(\sqrt{N})$. Here trivial solutions are ones of the form (x,y,y,x,y) or (y,x,y,x,y)

Avoid this 5-var eqn ⇒ avoid $x_1 + x_2 = x_4 + x_5$, i.e., a Sidon set, thus $O(\sqrt{N})$ size

Theorem (CFSZ). The maximum size of a Sidon subset of $[N]$ without a solution in distinct variables to the equation

$$x_1 + x_2 + x_3 + x_4 = 4x_5$$

is at most $o(\sqrt{N})$ and at least $N^{1/2-o(1)}$.

Erdős–Simonovits compactness conjecture

Excluding a finite set of graphs ≈ excluding the worst one

Conjecture. Given graphs $F_1, \ldots, F_k$, $\exists i, c > 0$:
max # edges in an n-vertex graph avoiding all $F_1, \ldots, F_k$
$\geq c \cdot$ max # edges in an n-vertex graph avoiding F_i

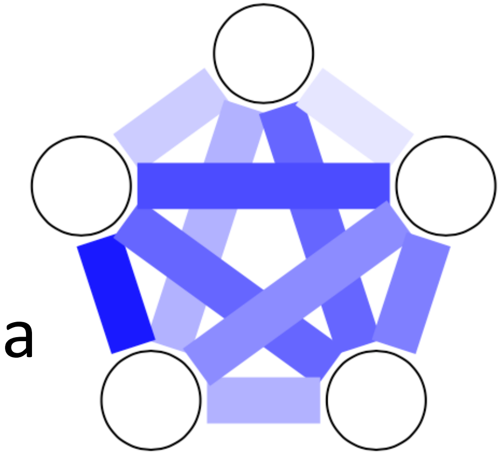
False for hypergraphs (due to Ruzsa–Szemerédi 6,3-theorem)

The equation-avoidance analog is false too! For subset of $[N]$

- Largest subset avoiding $x_1 + x_2 = x_3 + x_4$ has size $\sim\sqrt{N}$ (Sidon sets)
- Largest subset avoiding $x_1 + x_2 + x_3 + x_4 = 4x_5$ has size $N^{1-o(1)}$ (Behrend)

But! Avoiding both equations simultaneously $\Rightarrow$ size $= o(\sqrt{N})$

Regularity recipe



1. **Partition** the vertex set using (sparse) regularity lemma
2. **Clean** up the graph
   - Remove edges from irregular pairs and very sparse pairs
   - *(Only for sparse regularity)* Remove edges from extra dense pairs
3. **Count** subgraphs

Removing dense spots:

If $o(n^2)$ C_4's, then $o(n^{3/2})$ edges lie between too-dense parts.

C_5 counting lemma

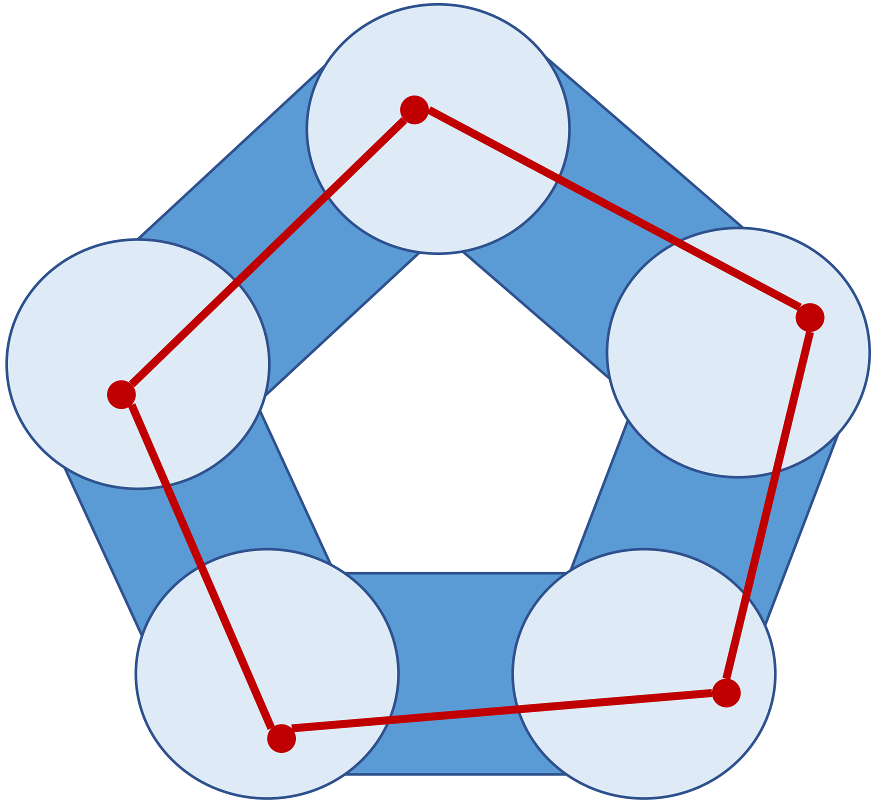
A **counting lemma** compares subgraph densities between two (weighted) graphs that are close in cut norm

C_5-counting lemma in graphs with not too many C_4's.

- G : 5-partite sparse graph with edge-density p
  - has $O(p^4 n^4)$ C_4's between adjacent parts
- G': is has edge-weights in $[0, Cp]$

If G and G' close in cut norm, then

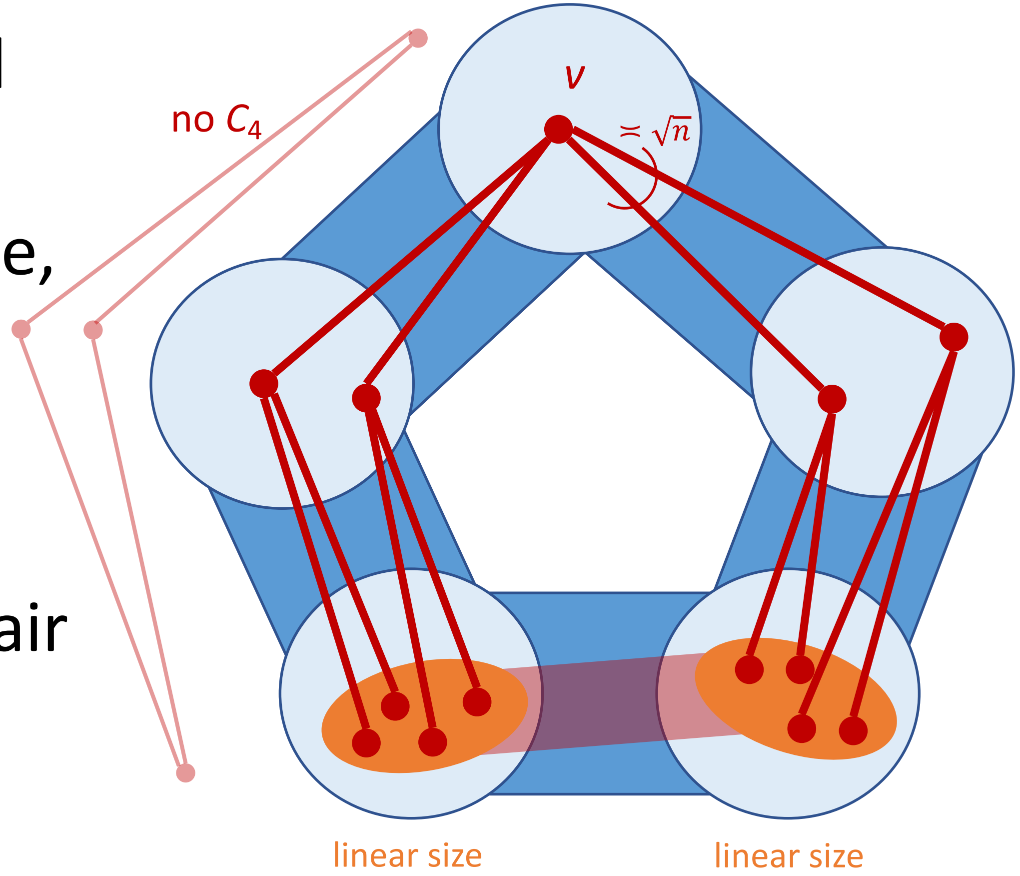
$$C_5\text{-density in } G > C_5\text{-density in } G' - o(p^5)$$



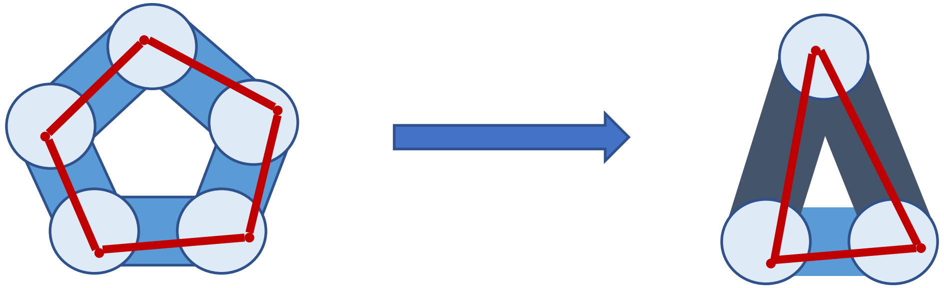
Being C_4-free helps counting C_5

A toy case: all vertex-degrees equal, and all bipartite graphs pseudorandom

Second neighborhood expands to linear size, thereby giving lots of C_5's



In general, analytic argument: replace two adjacent sparse pairs by a single "dense" pair



Proof of sparse removal lemma

Sparse $\{C_3, C_5\}$-removal lemma (CFSZ). Every n-vertex graph with

- $o(n^2)$ C_4's and
- $o(n^{5/2})$ C_5's

can be made C_5-free and C_3-free by removing $o(n^{3/2})$ edges.

1. **Partition.** Apply regularity partition to approximate G by a weighted graph G'
2. **Clean.** Remove $o(n^{3/2})$ edges from irregular, too-sparse, or too-dense pairs in G
3. **Count.** If any C_3 or C_5 remain in G, then can find C_5 in G'. Apply counting lemma to deduce that G has lots of C_5's

Proof of equation-avoidance in Sidon sets

Theorem (CFSZ). If $A \subset [N]$ is a Sidon set without nontrivial solution to

$$x_1 + x_2 + x_3 + x_4 = 4x_5,$$

then $|A| = o(\sqrt{N})$.

Set up a 5-partite graph

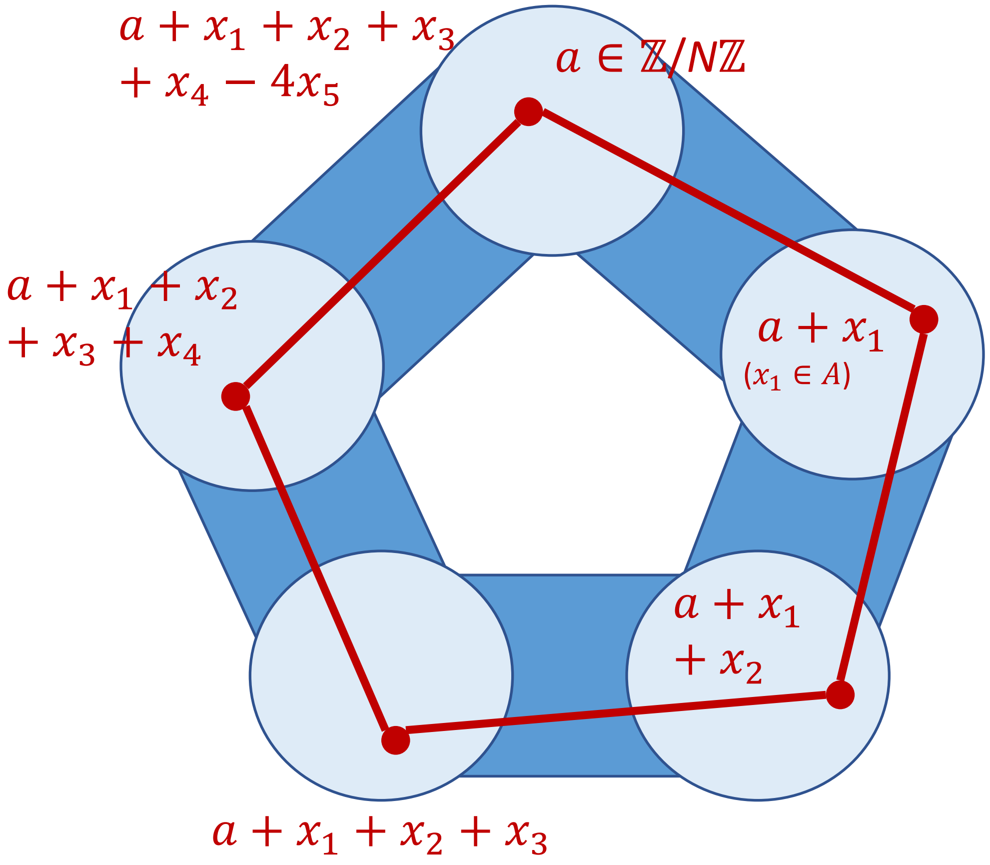
Avoiding $x_1 + x_2 + x_3 + x_4 = 4x_5$

$\Rightarrow$ every edge lies in exactly one C_5

Sidon $\Rightarrow$ C_4-free between parts

Sparse C_5-removal lemma $\Rightarrow o(N^{3/2})$ edges

$\Rightarrow |A| = o(\sqrt{N})$



each vertex set = $\mathbb{Z}/N\mathbb{Z}$

Sparse regularity and applications

- New **C_5-counting lemma** in sparse graphs with not too many C_4's
- **Sparse removal lemmas.** Every n-vertex graph with $o(n^2)$ C_4's and $o(n^{5/2})$ C_5's can be made C_5-free and C_3-free by removing $o(n^{3/2})$ edges.
- An n-vertex C_5-free graph can be made triangle-free by deleting $o(n^{3/2})$ edges.
  - Applications to extremal problems on hypergraphs
- A Sidon subset of $[N]$ avoiding solutions to a fixed 5-variable translation-invariant equation has size $o(\sqrt{N})$