**CS 59000 CTT Current Topics in Theoretical CS** Sep 11, 2012

# Lecture 7

*Lecturer: Elena Grigorescu* *Scribe: Greg Hurst*

## Today:

- We'll show how the $\triangle$-removal lemma (proved in the previous lecture) implies Roth's theorem.
- We'll show a lower bound on testing $\triangle$-freeness, due to [1].

First recall the $\triangle$-removal lemma.

**Lemma 1** *For all $\epsilon > 0$ there exists $\delta(\epsilon) > 0$ such that every graph $G$ on $n$ vertices which is $\epsilon$-far from $\triangle$-free contains $\delta\binom{n}{3}$ $\triangle$'s.*

Recall that $\epsilon$-far means we need to remove $\epsilon\binom{n}{2}$ edges to remove all $\triangle$'s from $G$.

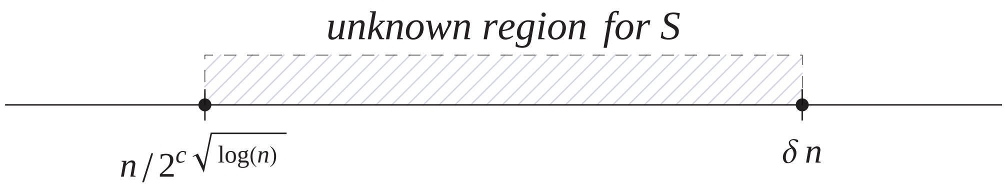
**Definition 2** *A* length-3-arithmetic progression *(or 3-AP) is a triple $\{x, x+\alpha, x+2\alpha\}$ for $x, \alpha \in \mathbb{Z}$, or equivalently, a tuple $\{a, b, c\}$ s.t. $a + c = 2b$, $a, b, c \in \mathbb{Z}$. The 3-AP is called* trivial *if $\alpha = 0$ (or equivalently, $a = b = c$), and it is called* non-trivial *otherwise.*

A question well-studied in the area of additive combinatorics is the following: what is the size of the largest set $S \subset [n] = \{1, 2, \dots, n\}$ that does not contain any 3-AP?

An example of such $S \subset [10]$ is $\{1, 2, 4, 5, 10\}$.

In this lecture we describe some known upper and lower bounds for $S$, namely

- upper bound on $S$: any $S$ with $|S| > \delta n$ must contain a 3-AP (Roth's theorem)
- lower bound on $S$: there exists $S$ with $|S| = \frac{n}{2^{c\sqrt{\log n}}} = \Omega(n^{1-\epsilon})$ that does not contain any 3-AP.(Behrend's theorem). We note that this result was shown in the 40's and has only recently been improved [2].



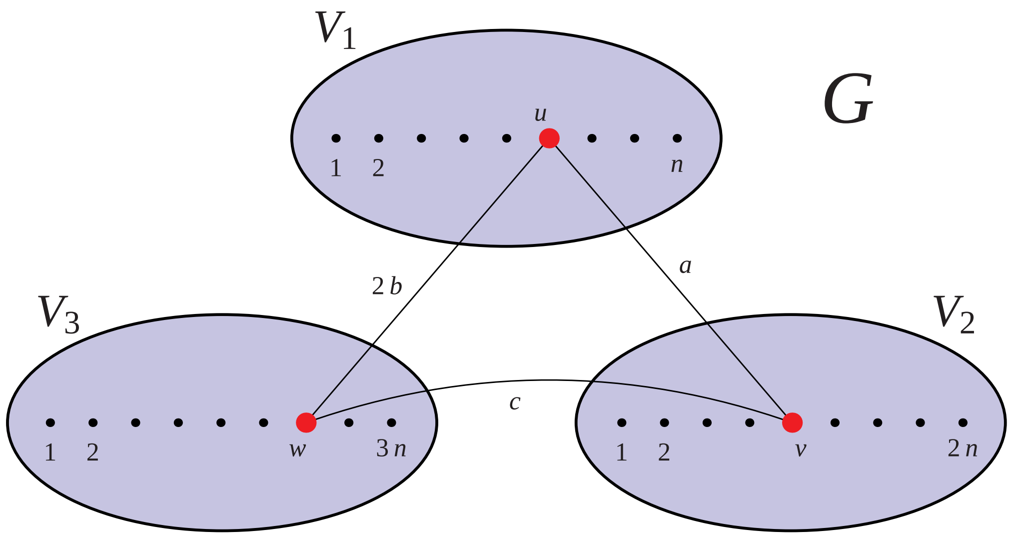

**Theorem 3 (Roth)** *For all $\epsilon > 0$ there exists $n_0(\epsilon) > 0$ such that for all $n > n_0(\epsilon)$, any $A \subset [n]$ with $|A| > \epsilon n$ must contain a 3-AP.*

**Proof**

Construct a graph $G = (V = V_1 \sqcup V_2 \sqcup V_3, E)$ where $V_1$ has $n$ vertices, $V_2$ has $2n$ vertices, $V_3$ has $3n$ vertices, and has edges that satisfy exactly one of the following:

- for $u \in V_1$ and $v \in V_2$, add edge $(u, v)$ is there is an $a \in A$ such that $v = u + a$

- for $v \in V_2$ and $w \in V_3$, add edge $(v, w)$ is there is an $c \in A$ such that $w = v + c$
- for $u \in V_1$ and $w \in V_3$, add edge $(u, w)$ is there is an $b \in A$ such that $w = u + 2b$




For $G$ we get $|V| = 6n$ and $|E| = n|A| + 2n|A| + n|A| = 4n|A|$.

Now every $\triangle$ in $G$ corresponds to a 3-AP in $A$ since $u+2b = w = v+c = u+a+c \implies 2b = a+c$. This means

$$\#\triangle\text{'s in } G = n \cdot (\#\text{non-trivial 3-AP's} + \#\text{trivial 3-AP}\,\triangle\text{'s}) = n \cdot (\#\text{non-trivial 3-AP's} + |A|).$$

Also

$$\#\text{edge disjoint } \triangle\text{'s} \geq |\{(x, x+a, x+2a) \mid x \in [n], a \in A\}| = n|A| \geq \epsilon n^2.$$

Since we need to remove at least one edge from the edge-disjoint triangles in order to distroy all the $\triangle$'s in $G$, this implies that

$$\text{dist}(G, \triangle\text{-freeness}) \geq \frac{\#\text{edge disjoint } \triangle\text{'s}}{\binom{n}{2}} \geq \frac{\epsilon}{4}$$

and so by the $\triangle$-removal lemma, $G$ has at least $\delta(\epsilon)n^3$ many triangles.

From here we conclude $n \cdot (\#\text{non-trivial 3-AP's} + |A|) \geq \delta n^3 \implies \#\text{non-trivial 3-AP's} > \delta n^2 - |A| > \delta n^2 - n \gg 1$.

■

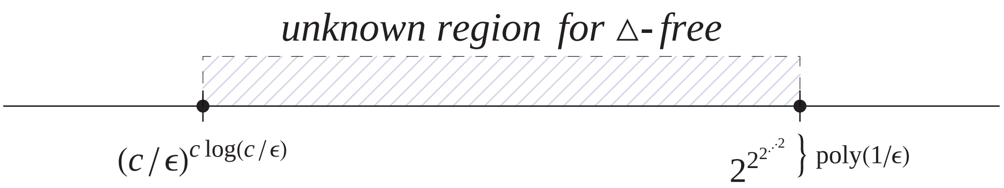
**Theorem 4 (Behrend)** *There exists $A \subset [n]$ of size $\frac{n}{2^{c\sqrt{\log n}}}$ such that $A$ has no 3-AP.*

We won't show this theorem here but we will show a nice connection to property testing, namely we will show how it implies a lower bound on the query complexity of tests for $\triangle$-freeness.

**Theorem 5 (Lower bound for testing $\triangle$-freeness)** *For all $\epsilon > 0$ there exists a graph $G$ on $n$ vertices such that any one sided error test must make $\Omega\left(\left(\frac{c}{\epsilon}\right)^{c\log(c/\epsilon)}\right)$ queries.*

The lower bound is shown by exhibiting a graph that is far from being $\triangle$-free yet does not have that many triangles (but it still has a constant fraction of the total number of triangles possible). In particular, the inverse of the fraction of triangles is superpolynomial in $\frac{1}{\epsilon}$, but much much smaller than the tower of exponentials upper bound from the triangle removal lemma (which in turn comes from the constant in Szemeredi's regularity lemma).

**Theorem 6 (Alon [1])** *There exists $c > 0$ such that for all $\epsilon > 0$ and $n > n_0(\epsilon)$ there exists a graph $G$ on $n$ vertices such that $G$ is $\epsilon$-far from $\triangle$-free and $G$ contains at most $\left(\frac{\epsilon}{c}\right)^{c\log(c/\epsilon)} \cdot \binom{n}{3}$ triangles..*




**Proof** [of Theorem 5] Let $G$ be the graph given by Theorem 6.

As we've seen before, one possible test could be: pick three vertices uniformly at random and check if they form a $\triangle$. If so reject. Then it follows that

$$P[\text{rej } G] = \frac{\#\triangle' s}{\binom{n}{3}} \leq \left(\frac{\epsilon}{c}\right)^{c\log(c/\epsilon)}.$$

Therefore, in order to reject $G$ w.p. $2/3$ we need to repeat this test at least $\frac{1}{100}\left(\frac{\epsilon}{c}\right)^{c\log(c/\epsilon)}$ many times.

Note that this does not complete the proof of the theorem sice we only analyzed one possible test, while the theorem states that any test must have this query complexity. Notice that the test above was non-adaptive and it only operated by checking the existence of triangles. One can imagine more interesting adaptive tests that might not need as many queries.

It turns out that in fact any (single sided) test (even fancy adaptive ones) can be reduced to the test above which simply picks a few vertices and makes a decision based on querying the edges spanned by these vertices. Importantly, the query complexity only blows up quadratically. We next state the formal theorem due to Goldreich and Trevisan [3] describing this result.

**Theorem 7** *If $T$ is a test (possibly adaptive) for a graph property $P$ that makes $q = q(\epsilon, n)$ queries then there exists a test $T'$ that, on input a graph $G$, picks $2q$ uniformly random vertices and accepts/rejects based on queries on the edges spanned by these vertices. The query complexity of $T'$ is thus $2q^2$. Moreover, if $T$ is single-sided so is $T'$. Such a test is called a* canonical test*.*

So the test shown above is a canonical test for $\triangle$-freeness. Therefore, to finish the proof of the lower bound for triangle freeness, note that if there existed a test with query complexity $q = o\left(\left(\frac{c}{\epsilon}\right)^{c\log(c/\epsilon)}\right) = \left(\frac{c}{\epsilon}\right)^{o(\log(1/\epsilon))}$, then there would exist a canonical test with query complexity $q^2 = \left(\frac{c}{\epsilon}\right)^{2\cdot o(\log(1/\epsilon))}$, a contradiction to the lower bound shown above for the canonical test.
■

**Proof of Theorem 6**
Consider the graph $G = G(V, E(G))$ from the proof of Roth's theorem, where the set $A \in [n]$ is this time the set given by Behrend's theorem.
For this proof we will switch the letter $n$ (the number of vertices) for the letter $m$.

Recalling facts from before, we have

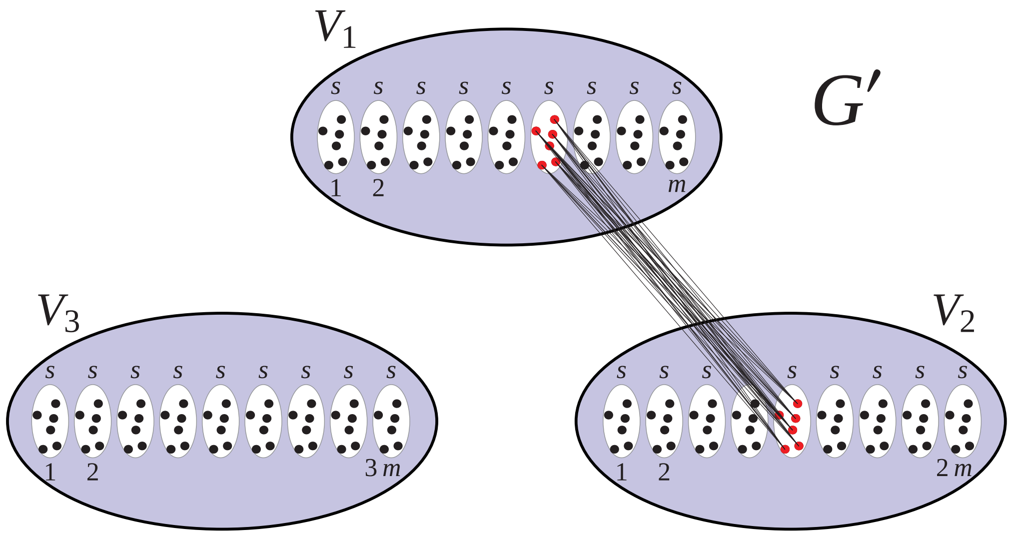
1. $|A| = \dfrac{m}{2^{c\sqrt{\log m}}}$
2. $|V| = 6m$,
3. $|E| = 4m|A|$,
4. #edge disjoint $\triangle$'s $= m|A|$, (since $A$ is triangle free, so the only possible triangles in $G$ are those of the form $\{x, x+a, x+2a\}$, with $x \in [m]$ and $a \in A$.
5. $\text{dist}(G, \triangle\text{-freeness}) \geq \frac{m|A|}{\binom{m}{2}} \geq \frac{1}{36}\frac{|A|}{m}$.

We want $G$ to be $\epsilon$-far from $\triangle$-free, so we want $m$ s.t. $\frac{1}{36}\frac{|A|}{m} > \epsilon$.
So choose $m$ such that

$$\left\lfloor \frac{1}{36}\frac{|A|}{m} \right\rfloor - 1 \leq m < \left\lfloor \frac{1}{36}\frac{|A|}{m} \right\rfloor \implies m = \Theta\left(2^{\log^2(c'/\epsilon)}\right) = \Theta\left(\left(\frac{c'}{\epsilon}\right)^{c'\log(c'/\epsilon)}\right),$$

for a suitable constant $c'$ (independent of $\epsilon$).
Now construct a graph $G'$ from $G$ as follows:

- for each $v \in G$, replace $v$ with $s = \Theta\left(\frac{n}{m}\right)$ vertices
- replace each edge with a complete bipartite graph (with $s$ vertices on the left and $s$ vertices on the right).




For $G'$ we get

1. $|V(G')| = 6ms = n$,
2. $|E(G')| = 4m|A|s^2$,
3. $\#\triangle$'s $= s^3(\#$ of $\triangle$'s in $G)$,
   and
4. $\#$ of disjoint $\triangle$'s $= s^2(\#$ of disjoint $\triangle$'s in $G)$.

So

$$\mathrm{dist}(G', \triangle\text{-freeness}) = \frac{\#\text{ disjoint }\triangle\text{'s}}{\binom{m}{2}} > \epsilon$$

(the last inequality simply follows from the choice of $m$ that we made).

This implies $G'$ is $\epsilon$-far from $\triangle$-free. Plugging in $s$, $m$ as functions of $\epsilon, n$ in item 3 above gives us that

$$\#\triangle\text{'s} < \Theta\left(\left(\frac{\epsilon}{c}\right)^{c\log(c/\epsilon)}\right)\binom{n}{3},$$

for some appropriately chosen $c$, as claimed in the theorem's statement.

■

## References


[1] Noga Alon. Testing subgraphs in large graphs. *Random Struct. Algorithms*, 21(3-4):359–370, 2002.

[2] Michael Elkin. An improved construction of progression-free sets. In *SODA*, pages 886–905, 2010.

[3] Oded Goldreich and Luca Trevisan. Three theorems regarding testing graph properties. *Random Struct. Algorithms*, 23(1):23–57, 2003.