Sublinear Algorithms		February 16, 2012
	Lecture 12	
Lecturer: Sofya Raskhodnikova		*Scribe(s): Dragos Nistor*

1 Introduction

This lecture contains proofs related to triangle-freeness.

Theorem 1 (Alon's Theorem)**.** *Every 1-sided-error ϵ-tester for triangle-freeness in the adjacency matrix model should make at least $\left(\frac{c}{e}\right)^{c\log\left(\frac{c}{e}\right)}$ queries.*

So we must construct graphs that are ϵ-far from triangle-free but have $o(\left(\frac{c}{e}\right)^{c\log\left(\frac{c}{e}\right)})$ triangles.

We will use the fact that graph properties are independent of vertex permutations to construct a special graph so that we can prove the lower bound on the queries needed.

Exercise 12.1 (Takehome exercise)**.** *Why does the complexity blow up quadratically in a canonical tester?*

2 Behrend's Theorem

Theorem 2 (Behrend's Theorem)**.** *For all m there exists a set $X \subseteq \{1, 2, ...m\}$ s.t. $|X| \geq \frac{m}{e^{10\sqrt{\log(m)}}}$ and the only solution to $x_1 + x_2 = 2x_3$ for $x_1, x_2, x_3 \in X$ is $x_1 = x_2 = x_3$.*

We will use some of the ideas presented by Behrend[Beh46]. The best known bound is presented by Elkin[Elk10].

Exercise 12.2 (Takehome exercise)**.** *Is it sufficient to have 2^4 instead of $e^{10\sqrt{\log(m)}}$?*

Proof of Theorem 2: Define many sets X_B, where

$$X_B = \{\sum_{i=0}^{k} x_i d^i \text{ such that } 0 \leq x_i < \frac{d}{2} \text{ and } B = \sum_{i=0}^{k} x^2\}$$

where d is an integer parameter and $k = \frac{\log m}{\log d} - 1$. This is a d-nary system where values fall into sets X_B. Additionally, all x_i are integers and $B > 0$.

Claim 3. *$X_B \subseteq \{1, 2, ...m\}$ for all B.*

Proof of Claim 3: We will prove that for every $x \in X_B, x \leq m$.

$$\begin{aligned} x &\leq d^{k+1}, \text{ } d\text{-nary system} \\ d^{k+1} &= d^{\frac{\log m}{\log d}}, \text{ definition of } k \\ d^{\frac{\log m}{\log d}} &= d^{\log_d m}, \text{ log base change} \\ d^{\log_d m} &= m \end{aligned}$$

□

Claim 4. *For all B, the only solution to $x + y = 2z$ for $x, y, z \in X_B$ is $x = y = z$.*

Proof of Claim 4: Suppose $x + y = 2z$ for some $x, y, z \in X_B$. Then, representing x, y, z in base d, we have that

$$\sum_{i=0}^{k} x_i d^i + \sum_{i=0}^{k} y_i d^i \quad = \quad \sum_{i=0}^{k} z_i d^i.$$

Because $x_i < \frac{d}{2}, y_i < \frac{d}{2}$, and $z_i < \frac{d}{2}$, there are no carries, and therefore no overflow. Since there are no carries, it must be that for all $i, x_i + y_i = 2z_i$. Consider the following convex function: $f(a) = a^2$. Since f is convex, we can apply Jensen's inequality to

$$\sum_{i=0}^{k} \frac{f(a_i)^2}{n}$$

and find that

$$\sum_{i=0}^{k} \frac{f(a_i)^2}{n} \quad \geq \quad f\left(\frac{1}{n}\sum_{i=0}^{k} a_i^2\right)$$

with equality holding only when all a_i are the same. Therefore, $\frac{x_i^2+y_i^2}{2} \geq \left(\frac{2z_i}{2}\right)^2$ with equality holding only when $x_i = y_i = z_i$. So, we sum over all i, and get that

$$\sum_{i=0}^{k} x_i + \sum_{i=0}^{k} y_i \quad \geq \quad \sum_{i=0}^{k} z_i$$

with equality holding only when $x_i = y_i = z_i$ for all i. But, equality must hold, because

$$\sum_{i=0}^{k} x_i = \sum_{i=0}^{k} y_i = \sum_{i=0}^{k} z_i = B.$$

Therefore, $x = y = z$ by the definition of $x, y, z \in X_B$. □

Claim 5. *For some B,* $|X_B| \geq \frac{m}{e^{10\sqrt{\log m}}}$.

Proof of Claim 5: $1 \leq B \leq (k+1)\left(\frac{d}{2}\right)^2$ since there are $k+1$ digits that we are summing up. B is also an integer, so there are at most $(k+1)\left(\frac{d}{2}\right)^2$ distinct sets X_B. Additionally, the number of integers covered by all sets has an upper bound of m and a lower bound of $\left(\frac{d}{2}\right)^{k+1} - 1 \geq \left(\frac{d}{2}\right)^k$ because of the restriction we placed on each x_i and B. Therefore, the size of some X_B is at least

$$\begin{aligned} \frac{\left(\frac{d}{2}\right)^k}{(k+1)\left(\frac{d}{2}\right)^2} &= \frac{\left(\frac{d}{2}\right)^{k-2}}{k+1} \\ &= \frac{d^{k+1}}{(k+1)(2^{k-2})(d^3)}. \end{aligned}$$

If we set $d = 2^{\sqrt{\log m}}, k = \sqrt{\log m} - 1$. Therefore,

$$\begin{aligned} |X_B| &\geq \frac{d^{k+1}}{(k+1)(2^{k-2})(d^3)} \\ &= \frac{m}{\sqrt{\log m}}\left(\frac{1}{2^{4\sqrt{\log m}}}\right). \end{aligned}$$

□

3 Initial Graph Construction

Now let's actually construct our special graph. Consider $V_1 = \{1, 2, ...m\}, V_2 = \{1, 2, ...2m\}, V_3 = \{1, 2, ...3m\}$. Add edges as follows: For each $a \in V_1, b \in V_2, c \in V_3$, for each $x \in X$, put an edge (a, b) if $b = a + x, (b, c)$ if $c = b + x$, and (a, c) if $c = a + 2x$. So, our intended triangles are $(i, i + x, i + 2x)$ for $i \in \{1, 2, ...m\}, x \in X$. There cannot be any other triangles in this graph because $x + y = 2z$ if and only if $x = y = z$, by construction. Additionally, this also means that all triangles are edge-disjoint. See Figure 1 for an example, where $m = 3$ and $X = \{2\}$. So, $|V| = 6m, |E| =$

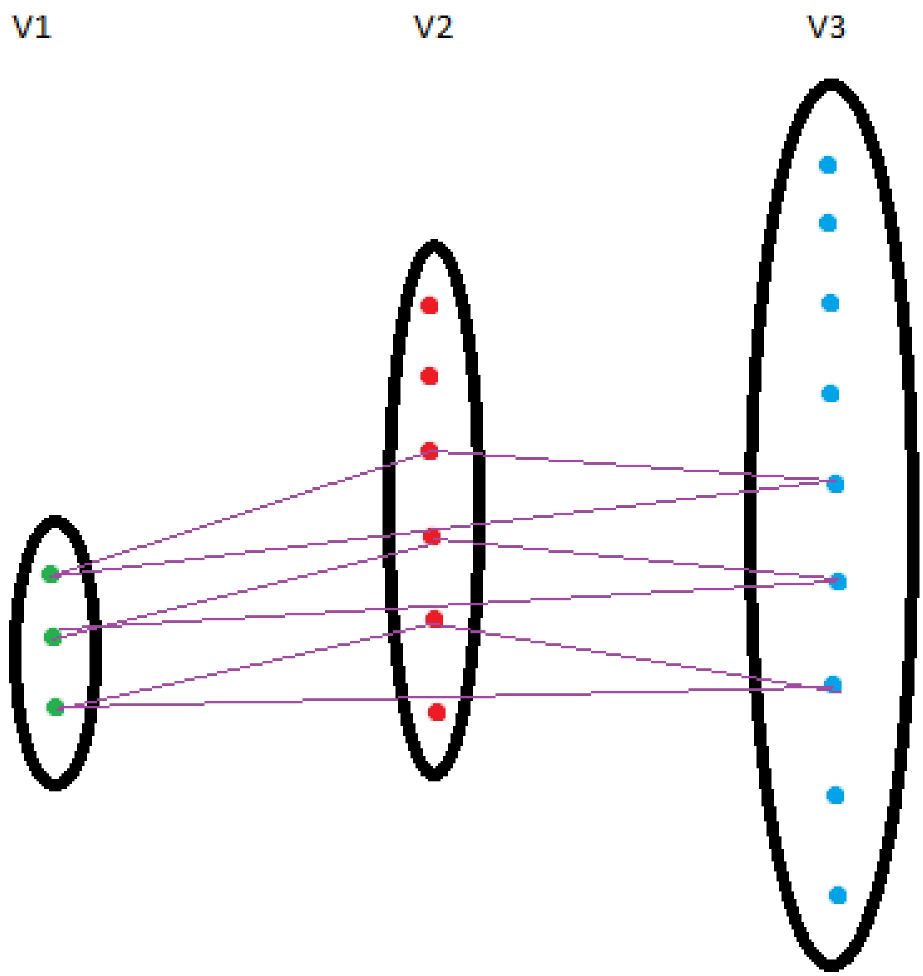

Figure 1: m = 3, X = {2}.

$3|X|m$, and # of edge-disjoint triangles $= |X|m$, since all triangles are edge-disjoint. We define

$$\begin{aligned} dist(\text{Graph, triangle-free}) &= \frac{\text{triangle count}}{n^2} \\ &= \frac{\text{triangle count}}{36m^2} \\ &= \frac{|X|m}{36m^2} = \frac{|X|m}{36m} \end{aligned}$$

which is unfortunately not a constant value. So, let's modify the construction of our graph by adding gadgets.

4 Modified Graph Construction

Replace each node v_i with a gadget v_i containing cluster of s nodes. If there was an edge (v_i, v_j) in our previous graph, create an edge between all nodes in gadget v_i and all nodes in gadget v_j. In this new graph, $|V| = 6ms, |E| = 3|X|ms^2$, triangle count $= \Theta(m|X|s^3)$ because for each 3 vertices, there are now $\binom{s}{3}$ triangles, and disjoint-edge triangles $= \Theta(m|X|s^2)$ because for each 3 vertices, there are s choices for the first, s choices for the second, and only one possibility for the third, otherwise they are not edge-disjoint. So now, given ϵ and n, we just pick m to be the largest integer satisfying $\frac{1}{m} \geq \left(\frac{c}{e}\right)^{c \log\left(\frac{c}{e}\right)}$. □

References

[Beh46] F. A. Behrend. On sets of integers which contain no three terms in arithmetic progression. *PNAS*, 32:331–332, 1946.

[Elk10] M. Elkin. An improved construction of progression-free sets. In *SODA*, pages 886–905, 2010.