6.896 Sublinear Time Algorithms	October 12, 2004

Lecture 9

Lecturer: Ronitt Rubinfeld | *Scribe: Akshay Patil*

In this lecture we will show that any tester for the property of being triangle free must have exponential dependence on ϵ.

Theorem 1 (Alon) *Let P_Δ = Property of not having Δ as a subgraph. $\exists c$ such that any 1-sided tester for P_Δ (in adjacency matrix model) needs $\geq (\frac{c}{\epsilon})^{c\log(\frac{c}{\epsilon})}$ queries.*

1 Some tools

1.1 A structural theorem

Theorem 2 (Goldreich Trevisan) *Let P be any graph property in an adjacency matrix model. Suppose T is a tester for P with query complexity $q(n,\epsilon)$. Then there is a tester T' for P of the following form:*

1. *select $2q(n,\epsilon)$ random nodes*
2. *query all edges*
3. *decide*

In addition, if T has 1-sided error, so does T'

Thus it is enough to show a lower bound on algorithms which follow the format of T' and have 1-sided error.

1.2 An additive #-theory lemma

Lemma 3 *$\forall m$, $\exists x \subset \{1...m\}$ s.t. $|x| \geq \frac{m}{e^{10\sqrt{\log m}}}$ and x has no nontrivial solution to the equation $x_1 + x_2 = 2x_3$ (i.e. the only solutions are of the form $x_1 = x_2 = x_3$)*

Proof [Theorem 2]
Let $d \in \mathbb{Z}$ (we'll choose the value later).
Let $k \leftarrow \lfloor \frac{\log m}{\log d} \rfloor - 1$
Define X_B as follows:

$$X_B = \left\{ \sum_{i=0}^{k} x_i d^i \,\middle|\, 0 \leq x_i \leq \frac{d}{2} \,;\, \sum_{i=0}^{k} x_i^2 = B \right\}$$

We can view $x \in X_B$ as a number written in base d

- $x_i's \leftrightarrow$ digits of x
- $x_i's$ small $\leftrightarrow$ summing any pair will not generate carries

To prove the theorem, pick $x \leftarrow X_B$ s.t. $|X_B|$ is maximized.
We need to show

1. $\forall B\ X_B \subseteq \{1...m\}$

   Proof $d^{k+1} \leq d^{(\lfloor \frac{\log m}{\log d} \rfloor - 1)+1} \leq m$ ■

2. Sum property holds

   Proof Say

$$\begin{aligned} x+y=2z \quad &\Leftrightarrow \quad \sum_{i=0}^{k} x_i d^i + \sum_{i=0}^{k} y_i d^i = 2\sum_{i=0}^{k} z_i d^i \\ &\Leftrightarrow \quad x_0+y_0=2z_0 \text{ (because there are no carries)} \\ &and \quad x_1+y_1=2z_1 \\ &\ldots \\ &and \quad x_k+y_k=2z_k \end{aligned}$$

   Since $f(a)=a^2$ is a convex function, we can use Jensen's $\neq$:

$$\frac{\sum f(a_i)}{n} \geq f(\frac{\sum a_i}{n})$$

   with equality holding only if the a_i's are all equal. This tells us that $\frac{x_i^2+y_i^2}{2} \geq (\frac{2z_i}{2})^2 \Rightarrow x_i^2+y_i^2 \geq 2z_i^2$ with equality only if $x_i=y_i=z_i$.

   If $x,y,z \in X_B$ then $\sum x_i^2 = \sum y_i^2 = \sum z_i^2 = B$.

   If $\neg(x=y=z)$ then $\exists i$ for which $\neg(x_i=y_i=z_i) \Rightarrow x_i^2+y_i^2 > 2z_i^2$ and for all other i, $x_i^2+y_i^2 \geq 2z_i^2 \Rightarrow \sum x_i^2 + \sum y_i^2 > \sum z_i^2 \Rightarrow B+B > 2B$ **Contradiction!**

   $\Rightarrow \nexists i \Rightarrow x=y=z$. ∎

3. $\exists B$ s.t. $|X_B| \geq \frac{m}{e^{10\sqrt{\log m}}}$

   Proof Well,

$$\begin{aligned} 0 \leq B \leq (k+1)(\frac{d}{2})^2 \quad &< kd^2 \\ |\cup X_B| \geq (\frac{d}{2})^{k+1} \quad &\geq (\tfrac{d}{2})^k \end{aligned}$$

   Therefore, some X_B must be of size $> \frac{(\frac{d}{2})^k}{kd^2}$. Using $d=e^{10\sqrt{\log m}}$ we get $k=\frac{\log m}{\log d} \approx \frac{\log m}{c\sqrt{\log m}} \approx c'\sqrt{\log m}$ and it all works out magically. ∎

∎

2 Construction of the lower bound

Consider graph G with sub-graphs $V^{(1)}, V^{(2)}, V^{(3)}$ of sizes $m, 2m$, and $3m$ respectively. Within each sub-graph there are no edges, but between the subgraphs there are $\Theta(m \cdot |x|)$ edges. For each node $i^{(1)}$ in $V^{(1)}$, $j^{(2)}$ in $V^{(2)}$ and $k^{(3)}$ in $V^{(3)}$, and for each $x \in X$ (the set from the #-theory lemma), G contains the edges $(i^{(1)},(i+x)^{(2)}),\ (j^{(2)},(j+x)^{(3)}),\ (k^{(1)},(k+2x)^{(3)})$.

G has the *intended cycles* $i^{(1)} \to (i+x)^2 \to (i+2x)^3$, of which there are $\Theta(m \cdot |x|)$.

Claim: The only Δs in G are the intended Δs (of which there are $\Theta(m \cdot |x|)$).

Proof Consider a cycle in the graph, $i^{(1)} \to j^{(2)} \to k^{(3)}$, this means that $\exists i,j$ and $x_1,x_2,x_3 \in X$ s.t. $k=i+2x_3=j+x_2=i+x_1+x_2 \Rightarrow x_1+x_2=2x_3 \Rightarrow x_1=x_2=x_3$ by the #-theory lemma. ∎

Claim: The intended Δs are edge-disjoint.

Proof A shared edge between $u^{(2)}$ and $v^{(3)}$ implies that there is an i, j, k, k' such that $i+k = j+k'$ and $i+2k = j+2k'$ which implies $k = k'$ so $i = j$. Similiar math shows no other edge is shared as well. ■

Since you need to delete as one edge to break each Δ, this means distance of the graph from Δ-free is $\Theta(m \cdot |x|)$.

Almost there. $|x|$ is only $\geq \frac{m}{e^{10\sqrt{\log m}}}$ and we need to be $\Theta(\epsilon m^2)$ edges away from Δ free.

Fix: Do an s-blowup. Each node in the old graph becomes an independent set of size s in the new graph. Similarly, each edge in the old graph becomes a complete bipartite graph on the s nodes. This means that:

$$\begin{aligned} \#\text{nodes} &\approx \Theta(ms) \\ \#\text{edges} &\approx \Theta(m|x|s^2) \\ \#\Delta\text{s} &\approx \Theta(m|x|s^3) \end{aligned}$$

Claim: There are $\geq \Theta(m|x|s^2)$ edge disjoint Δs

Proof To appear in the homework. ■

Given ϵ, pick m to be the largest integer satisfying the equation $\epsilon \leq \frac{1}{3^8 e^{10\sqrt{\log m}}}$. Thus $m \geq (\frac{c}{\epsilon})^{c \log \frac{c}{\epsilon}}$

Choose $s = \lfloor \frac{2n}{6m} \rfloor$. This means that #edges $\sim$ distance $\sim \epsilon n^2$ and $\#\Delta$s $\leq \frac{\epsilon^{c \log(\frac{c}{\epsilon})}}{c} m^3$.

Because of the structural lemma, we assume that the algorithm has 1-sided error and samples a random subset of the vertices. Since it has 1-sided error, it can only fail if it actually finds a triangle. If we take a sample of size q, E[# Δs in sample] $\leq (\frac{q}{3})(\frac{\epsilon^{c \log \frac{c}{\epsilon}}}{c})n^3 << 1$ unless $q > (\frac{c'}{\epsilon})^{c' \log \frac{c'}{\epsilon}}$ for some c'.