6.896 Sublinear Time Algorithms	April 10, 2007
Lecture 17	
Lecturer: Ronitt Rubinfeld	*Scribe: Jacob Scott*

1 Alon's Theorem (adjacency matrix model)

Theorem 1 *Let H be a fixed graph with $|H| = h$, and P_H be the property of not having H as a subgraph. Then,*

1. *If H is bipartite:*
   (a) *$\exists$ a 2-sided tester for $P(H)$ with $O(1/\epsilon)$ queries*
   (b) *$\exists$ a 1-sided tester for $P(H)$ with $O(h^2(1/\epsilon)^{h^2/4})$ queries*
2. *If H is not bipartite, $\exists c$ such that any $1-$sided error tester needs $\geq (\frac{c}{\epsilon})^{c\log(1/\epsilon)}$*

We will prove (2) for triangles, that is, that testing triangle-freeness with 1-sided error requires queries superpolynomial in ϵ.

2 Structure of the Algorithm (Goldreich Trevisan)

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

1. *Select a random subset of $2q(n, \epsilon)$ nodes*
2. *Query all pairs in subset*
3. *Flip coins and make a decision*

If T has 1-sided error, then so does T'.

We note the following important properties of T':

1. T' has query complexity $O((q(n, \epsilon))^2)$
2. T' is non-adaptive
3. If, and only if, T' cannot decide P in $O(q''(n, \epsilon))$ queries, then T cannot decide P in $O(\sqrt{q''(n, \epsilon)}$ queries

Thus, a superpolynomial lower bound for T' implies a superpolynomial bound for T. We later show lower bounds for such a T' and by Goldreich-Trevisan can conclude that our bounds apply to all testers.

3 Number Theory Lemma

Lemma 3 $\forall m, \exists X \subset M = \{1, 2, \ldots, m\}$ *of size at least* $|X| > e^{m/(10\sqrt{\log m})}$*, with no nontrivial solution to* $X_1 + X_2 = 2X_3$ *(i.e., the only solution is* $X_1 = X_2 = X_3$*).*

Proof:
Let d be an integer, and $k = \lfloor \frac{\log m}{\log d} \rfloor - 1$.

Definition 1 *Let* $W_B = \left\{ \sum_{i=0}^{k} X_i d^i | \forall i \;\, X_i < d/2, \wedge \sum X_i^2 = B \right\}$, *where* B *is chosen to maximize* $|W_B|$.

Consider $x \in W_B$ as a base d vector, $x = [x_0, x_1, \ldots, x_k]$. Note that adding together two elements in W_B will result in a vector whose maximum value is $d-1$, thus not causing any 'carries'. Furthermore, $\max_x x \in W_B \leq \sum_{i=0}^{k} \frac{d}{2} d^i \leq d^{k+1} \leq d^{\lfloor \frac{\log m}{\log d} \rfloor} \leq d^{\log_d m} \leq m$.

Fact 4 *The Sum Property states that given* $x, y, z \in W_B$, $x + y = 2z \Leftrightarrow x_i + y_i = 2z_i$ *for all* i.

Proof of Fact:

$$\begin{aligned} \forall q \in W_B, \sum_{i=1}^{k} q_i d^i &= B \\ x + y = 2z &\Leftrightarrow \sum_{i=1}^{k} x_i d^i + \sum_{i=1}^{k} y_i d^i = 2 \sum_{i=1}^{k} z_i d^i \\ &\Leftrightarrow \forall i, x_i + y_i = 2z_i \end{aligned}$$

Here the last step stems from addition not generating carries for any d^i, as mentioned above. ■

Claim 5 *If* $x_i + y_i = 2z_i$, *then* $x_i^2 + y_i^2 \geq 2z_i^2$, *with equality if and only if* $x_i = y_i = z_i$.

Proof of Claim: $f(a) = a^2$ is convex. By Jensen's inequality, $\sum_{i=1}^{n} f(a_i)/n \geq f(\sum a_i/n)$ with equality only if all a_i are identical. ■

Proof of Lemma:

First, we prove that all solutions to $x + y = 2z$ must be trivial. Since $\forall q \in W_B, \sum_{i=1}^{k} q_i d^i = B$, it must be the case that $\sum x_i^2 + \sum y_i^2 = 2B = 2 \sum z_i^2$. By the Sum Property and the claim, for each i, $x_i^2 + y_i^2 \geq 2z_i^2$. However, if $\neg(x = y = z)$ then $\exists i$ such that $\neg(x_i = y_i = z_i)$. By the claim, it must then be that $x_i^2 + y_i^2 > 2z_i^2$. But $\forall\ j \neq i$, $x_j^2 + y_j^2 \geq 2z_j^2$. Thus, if $\exists i$ such that $\neg(x_i = y_i = z_i)$, then $\sum_i x_i^2 + \sum_i y_i^2 > 2 \sum_i z_i^2$, which is a contradiction. Therefor, $\forall i, x_i = y_i = z_i$, $x = y = z$, and all solutions are trivial.

Finally, we must consider the size $|W_B|$ of W_B. From the definition, we know that $B \leq (k+1)(d/2)^2 \leq kd^2$. It is also the case that every vector v with entries $< d/2$ has $\sum_i v_i^2 = B_j$ for some B_j. Thus

$$\sum_B |W_B| = |\cup W_B| \geq (d/2)^{k+1} > (d/2)^k$$

By the pigeonhole principle, we conclude that $\exists j$ such that W_{B_j} is of size $\geq (d/2)^k/(kd^2)$.If we set $d = e^{10\sqrt{\log m}}$, then $k \approx \sqrt{\log m}/10$, and $|W_B| = |W_{B_j}| \geq e^{m/(10\sqrt{\log m})}$, which proves the Lemma. ■

4 Next Lecture

In the next lecture we will combine Goldreich-Trevisan with our number theory lemma to produce hard graphs for which any tester for 1-sided triangle freeness requires $\geq (\frac{c}{\epsilon})^{c \log(1/\epsilon)}$ queries.