6.896 Sublinear Time Algorithms	April 5, 2007
Lecture 16	
Lecturer: Ronitt Rubinfeld	*Scribe: Adriana Lopez*

Recall from last lecture:

Definition 1 *The* **density** *between two disjoint non-empty sets $A, B \subseteq V$ is given by the formula $d(A,B) = \frac{e(A,B)}{|A||B|}$, where $e(A,B)$ is the number of edges between A and B.*

Definition 2 *We say $A, B \subseteq V$ are* **γ-regular** *if $\forall A' \subseteq A, B' \subseteq B$ such that $|A'| \geq \gamma|A|, |B'| \geq \gamma|B|$, the following inequality holds: $|d(A', B') - d(A,B)| < \gamma$.*

Lemma 3 (Szemeredi's Regularity Lemma) *For all m and for any $\epsilon > 0$, there exists $T = T(m, \epsilon)$ such that if $G = (V, E)$ with $|V| > T$, and if A is an equipartition of V into m sets, then there exists an equipartition B, a refinement of A, with k sets such that $m \leq k \leq T$ and at most $\epsilon\binom{k}{2}$ set pairs are not ϵ-regular.*

This lemma says that any large enough graph can in some sense have a constant-size description that makes it look random with respect to the densities of set pairs. The constant T, however, is very large:

$$T \approx 2^{2^{2^{2^{\cdots}}}} \Big\}_{\frac{1}{\epsilon^5} \text{ times}}$$

We will use the regularity lemma to prove the existence of a constant-time tester for triangle freeness, and therefore the running time of the tester will just as large (constant, but large).

1 Testing Triangle Freeness

Recall the following lemma from last lecture. We will use this lemma, along with theorem 5, to create a constant-time tester for triangle freeness.

Lemma 4 *For all $\eta > 0$, there exist $\gamma^{\triangle} = \gamma^{\triangle}(\eta)$ and $\delta^{\triangle} = \delta^{\triangle}(\eta)$ such that if A, B, C are disjoint subsets of V such that each pair is $\gamma^{\triangle}$-regular and has density at least η, then G contains at least $\delta^{\triangle}|A||B||C|$ distinct triangles (each triangle having one vertex in each of A, B, C).*

Theorem 5 *For all ϵ, there exists $\delta = \delta(\epsilon)$ such that $\forall G = (V, E)$ such that $|V| = n$ and G is ϵ-far from triangle-free in the adjacency matrix model (ie. we must remove at least ϵn^2 edges to make G triangle-free), then G has at least $\delta\binom{n}{3}$ triangles.*

Corollary 6 *The exists a one-sided property tester T for triangle-freeness whose running time is independent of the size of the graph G.*

Proof Consider the following tester T:

- Do $O(\delta(\epsilon)^{-1})$ times:
  - Pick v_1, v_2, v_3 independently and uniformly at random from V.
  - If the chosen vertices form a triangle, reject and halt.
- If tester hasn't halted, pass.

If G is triangle-free then no matter what vertices are chosen, they will never form a triangle and thus, the T will output pass. If, however, G is ϵ-far from being triangle-free, then by theorem 5 we know that G has at least $\delta\binom{n}{3}$ distinct triangles. Thus,

$$\Pr[\text{pick a triangle in one round}] \geq \frac{\delta\binom{n}{3}}{\binom{n}{3}} \geq \delta$$

Let c be the constant implied by the expression $O(\delta(\epsilon)^{-1})$, so that

$$\Pr[T \text{ outputs pass}] = \Pr[\text{never see a triangle}] \leq (1-\delta)^{\frac{\epsilon}{\delta}} \tilde{e}^{-c} < \frac{1}{4}$$

(the last inequality follows from choosing c sufficiently large). ■

Proof of Theorem 5: Take an arbitrary partition of G into $\frac{5}{\epsilon}$ sets, and use Szemeredi's regularity lemma (lemma 3) to find a partition $\{V_1, V_2, \cdots, V_k\}$ such that $\frac{5}{\epsilon} \leq k \leq T(\frac{5}{\epsilon}, \epsilon')$ for $\epsilon' = \min\{\frac{\epsilon}{5}, \gamma^{\triangle}(\frac{\epsilon}{5})\}$ and such that at most $\epsilon' \binom{k}{2}$ pairs are not ϵ-regular. First notice that the number of nodes in each set is roughly $\frac{n}{k}$ (we will assume from now on that it is exactly $\frac{n}{k}$). Thus,

$$\frac{\epsilon n}{5} \geq \frac{n}{k} \geq \frac{n}{T(\frac{5}{\epsilon}, \epsilon')} \tag{1}$$

Now construct a graph G' from G as follows:

1. delete edges of G internal to any V_i
   (# edges deleted in step 1 $\leq n\frac{n}{k} \leq \frac{\epsilon n^2}{5}$)
2. delete edges between non ϵ'-regular pairs V_i, V_j
   (# edges deleted in step 2 $\leq \epsilon' \binom{k}{2} (\frac{n}{k})^2 \leq \frac{\epsilon' n^2}{2} \leq \frac{\epsilon n^2}{10}$. This is the reason why we want $\epsilon' \leq \frac{\epsilon}{5}$)
3. delete edges between pairs whose density is less than $\frac{\epsilon}{5}$ ("low density" pairs)
   (# edges deleted in step 3 $\leq \displaystyle\sum_{\substack{\text{low density}\\ \text{pairs}}} \left(\frac{\epsilon}{5}\right)\left(\frac{n}{k}\right)^2 \leq \left(\frac{\epsilon}{5}\right)\binom{n}{2}$. The last inequality follows from the fact that $\sum_{\text{all pairs}} (\frac{n}{k})^2 \leq \binom{n}{2}$).

The total number of edges we delete is therefore at most $\frac{3\epsilon n^2}{5} + (\frac{\epsilon}{5})\binom{n}{2} < \epsilon n^2$. Since G is ϵ-far from being triangle-free, this tells us that G' has at least one triangle. We will prove that in fact, G' has many triangles. Let x, y, z be the vertices of such triangle. Then there exist i, j, l such that $x \in V_i, y \in V_j, z \in V_l$. Furthermore, it must be the case that i, j, l are all distinct since we deleted all edges internal to any V_i (step 1). We also know that all possible pairs between V_i, V_j, V_l have density at least $\frac{\epsilon}{5}$ because we deleted all low density pairs (step 3), and they are ϵ'-regular because we deleted all non ϵ'-regular pairs (step 2). We can therefore use lemma 4 with $\eta = \frac{\epsilon}{5}$. Thus,

$$\#\text{ triangles in } G \geq \delta^{\triangle}\left(\frac{\epsilon}{5}\right)|V_i||V_j||V_l| \geq \delta^{\triangle}\left(\frac{n}{T(\frac{5}{\epsilon,\epsilon'})}\right)^3 \geq \delta' n^3$$

where δ' is a function of ϵ but not n (such function exists). The second inequality follows from equation 1. ■

We have shown that we can test triangle freeness in constant time. However, this constant is very large since it depends on the "tower" constant of the regularity lemma. It is still open whether we can test triangle freeness in "reasonable" constant time.

2 Testing in Constant Time

In the adjacency matrix model, what can be tested in constant time with one-sided error? The answer is all *hereditary graph properties*, those that are closed under vertex removal but not edge removal (although vertex removal implies the removal of all edges adjacent to it). Examples of these properties are:

- all monotone graph properties
  - triangle freeness
  - bipartiteness
  - planarity
  - k-colorability
- perfect graph
- interval graph

We also want to know, what can we test in constant time with two-sided error? In this case, we can test properties that can be reduced to one of finitely many Szemeredi-type partitions. For example, there exists a constant-time algorithm that answers the following question: "given a graph G, a constant c, and densities $\{d_{ij}\}$, is it possible to partition G into c parts $V_1, V_2, \ldots, V_c$ such that for $i \neq j$, $d(V_i, V_j) = d_{ij}$?