6.896 Sublinear Time Algorithms	April 3, 2007
Lecture 15	
Lecturer: Kevin Matulef	*Scribe: Huy Le Nguyen*

1 Definitions and Proof Plan

Lemma 1 *(Szemeredi's Regularity Lemma). For any ϵ, there exists C_ϵ such that any graph $G = (V, E)$ can be partitioned into sets $V_0, V_1, \ldots, V_k$ $k \leq C_\epsilon$ such that*

1. $|V_0| \leq \epsilon V$
2. $|V_1| = |V_2| = \ldots = |V_k|$
3. *All but at most $\epsilon\binom{k}{2}$ pairs (V_i, V_j) are ϵ-regular*

Definition 2 *The density between disjoint vertex sets A, B is $d(A, B) = \frac{e(A,B)}{|A|\cdot|B|}$. We also define $d(A, A) = \frac{2e(A,A)}{|A|^2}$.*

Definition 3 *A and B are ϵ-regular if $\forall A' \subseteq A, |A'| > \epsilon|A|, B' \subseteq B, |B'| > \epsilon|B|$, we have $|d(A', B') - d(A, B)| < \epsilon$.*

One application of Szemeredi's Regularity Lemma that we mentioned last time was testing triangle-freeness of graph. The algorithm is very simple, just taking random triples and testing if they form triangles. We used an edge removal argument to prove the correctness of the algorithm. After partitioning the graph into components by Szemeredi's Regularity Lemma, we cleaned it up by removing internal edges of the components and edges between non ϵ-regular pairs. Afterwards, since all pairs of components are ϵ-regular, we can restrict ourselves to testing ϵ-regular pairs, which is much easier because of its regularity.

To prove Szemeredi's Regularity Lemma, we will follow the following sketch:

1. Start with arbitrary equipartition $V_1, \ldots, V_t$
2. Define "disorder" function $f(V_0, \ldots, V_k) \in [0, 1]$.
3. If partition doesn't satisfy lemma, refine it. Show that refinement increases f by $\Omega(\epsilon^5)$
4. Repeat at most $O(\frac{1}{\epsilon^5})$ times.

Definition 4 *Let $p_i = \frac{|V_i|}{V}$. Let X be the density $d(V_i, V_j)$ between two randomly chosen vertices z, t ($z \in V_i, t \in V_j$). We define $f(V_0, \ldots, V_k) = E[X^2] = \sum_{i,j} p_i p_j d(V_i, V_j)^2$.*

We can compute $E[X]$ as follow:

$$\begin{aligned}
E[X] &= \sum_{i,j} p_i p_j d(V_i, V_j) \\
&= 2\sum_{i<j} p_i p_j d(V_i, V_j) + \sum_i p_i^2 d(V_i, V_i) \\
&= 2\sum_{i<j} \frac{|V_i|}{|V|} \cdot \frac{|V_j|}{|V|} \cdot \frac{e(V_i, V_j)}{|V_i|} \cdot |V_j| + \sum_i \frac{|V_i|^2}{|V|^2} \cdot \frac{2e(V_i, V_i)}{|V_i|^2} \\
&= \frac{2}{|V|^2} \sum_{i<j} e(V_i, V_j) + \frac{2}{|V|^2} \sum_i e(V_i, V_i) \\
&= \frac{2|E|}{|V|^2}
\end{aligned}$$

Because the value of E[X] is independent of the partition, the change in f reflects the change in the variance of X.

2 The Proof

Lemma 5 *We start with a partition of the graph into vertex sets $V, W_1, W_2, \ldots W_k$. Subdividing set V into sets A and $\overline{A}$ can not decrease f.*

Proof Notice that E[X] doesn't change:

$$p_V p_W d(V, W) = p_A p_W d(A, W) + p_{\overline{A}} p_W d(\overline{A}, W)$$

We define:

$$\begin{aligned}\mathbf{x} &= \langle \sqrt{p_A p_W}, \sqrt{p_{\overline{A}} p_W} \rangle \\ \mathbf{y} &= \langle \sqrt{p_A p_W} d(A, W), \sqrt{p_{\overline{A}} p_W} d(\overline{A}, W) \rangle\end{aligned}$$

By Cauchy-Schwarz inequality, we have:

$$\begin{aligned}(\mathbf{x} \cdot \mathbf{y})^2 &\leq |\mathbf{x}|^2 \cdot |\mathbf{y}|^2 \\ p_V p_W d(V, W)^2 &\leq p_A p_W d(A, W)^2 + p_{\overline{A}} p_W d(\overline{A}, W)^2\end{aligned}$$

■

Lemma 6 *Let V, W be a non-ϵ-regular pair. So $\exists A \subseteq V, |A| > \epsilon|V|, B \subseteq W, |B| > \epsilon|W|, |d(A, B) - d(V, W)| > \epsilon$. Subdivide V into A, $\overline{A}$ and W into B, $\overline{B}$. f increases by at least $p_V p_W \epsilon^4$.*

Proof WLOG assume that $d(A, B) \geq d(V, W) + \epsilon$. f increases the least when

$$\begin{aligned}|A| &= \epsilon|V| \\ |B| &= \epsilon|W| \\ d(A, B) &= d(V, W) + \epsilon\end{aligned}$$

Before splitting, the pair V, W contribute $p_V p_W d(V, W)^2$ to f. After splitting they contribute:

$$\begin{aligned}&p_A p_B d(A, B)^2 + p_A p_{\overline{B}} d(A, \overline{B}^2 + p_{\overline{A}} p_B d(\overline{A}, B)^2 + p_{\overline{A}} p_{\overline{B}} d(\overline{A}, \overline{B})^2 \\ =&(\epsilon p_V)(\epsilon p_W)(d(V, W) + \epsilon)^2 + (\epsilon p_V)(1 - \epsilon) p_W (\cdot)^2 + \ldots \\ =&p_V p_W d(V, W)^2 + p_V p_W \epsilon^4 + \ldots\end{aligned}$$

Thus, subdividing V and W increases the contribution of the pair by at least $p_V p_W \epsilon^4$. Also notice that by the previous lemma, the contribution of other pairs is not negatively affected by the subdivision of V and W. Therefore, overall, f increases by at least $p_V p_W \epsilon^4$. ■

Now we already have all the components needed for the proof of Szemeredi's Regularity Lemma.
Proof Start with an arbitrary partition P of t components. If P doesn't satisfy the lemma, we can refine P by fixing all non-ϵ-regular pairs using lemma 6. Splitting subset never decreases f so we can perform the splits of all non-ϵ-regular pairs at once and increase f by at least the total increases we get from each split. V doesn't satisfy the lemma so it has at least $\epsilon\binom{k}{2}$ non-ϵ-regular pairs. Because all subsets have the same size, splitting all pairs increases f by at least ϵ^5. A set V can be splitted t-1 times so it could be divided into 2^{t-1} subsets. After dividing, the new subsets might have different size so we

have to divide them further so that all set have the same size. Because there could be some remainder in each of the division, we have to collect all those remaining vertices and put them into V_0. There are not many remaining vertices so we can put them all into V_0 and still have $|V_0| < \epsilon|V|$. Because f is upper-bounded by 1 and every time we refine P, f increases by at least ϵ^5, we only have to refine $\frac{1}{\epsilon^5}$ times. Every time we refine P, we exponentiate the number of sets once so we exponentiate the original number of sets $\frac{1}{\epsilon^5}$ times but this is still a constant only depending on ϵ and the number of sets we start with. ■