Sublinear Algorithms	November 5&10, 2015
Lecture 21-22	
Lecturer: Sofya Raskhodnikova	*Scribe(s):*

1 Introduction

In this lectures, we state (but do not prove) the Regularity Lemma. We also present an algorithm to test for triangle-freeness in the dense graph model, which is a special case of the algorithm by Alon et al. [AFKS00]. The running time of this algorithm is independent of the size of the input graph; however, the dependence on $1/\epsilon$ is very large ($2^{2^{\cdot^{\cdot^{\cdot^{2}}}}}$, where the number of exponentiations is $1/\epsilon$). The result is of combinatorial interest, but the algorithm is not very practical. This line of work culminated in [AFNS09] that provided a combinatorial characterization of graph properties in the dense graph model that can be tested in time independent of the size of the graph. This characterization is also based on the Regularity Lemma.

2 Testing $\triangle$-Freeness and the Regularity Lemma

2.1 Testing $\triangle$-freeness

We look at a special case of a result from [AFKS00] for the property of $\triangle$-freeness. The more general analysis uses similar techniques.

Input: Parameter $\epsilon \in (0, 1)$; access to an undirected graph $G = (V, E)$ given by an adjacency matrix.

Goal: ϵ-tester for $\triangle$-freeness.

The Test: Repeat s times:

1. Randomly pick v_1, v_2, v_3 from V.
2. If v_1, v_2, v_3 form a triangle, reject.

How big should s be? The following theorem gives us the answer.

Theorem 1 ($\triangle$-removal lemma)**.** *$\forall \epsilon\ \exists \delta = \delta(\epsilon)$ such that if a graph G on n nodes is ϵ-far from $\triangle$-free then G contains $\geq \delta\binom{n}{3}$ distinct triangles.*

Letting $s = \Theta(\frac{1}{\delta})$ will give us the required ϵ-tester. Note that it is very easy to prove that if G is ϵ-far from $\triangle$-free, it contains at least ϵn^2 triangles. This theorem is asymptotically better than this easy observation — it gives us $\Theta(n^3)$ triangles for constant ϵ.

2.2 The Regularity Lemma

Before turning to the proof of Theorem 1, we must provide some background on the main tool used in the proof: Szemerédi's Regularity Lemma [Sze78]. We will first need a few definitions.

Definition 2 (Density)**.** *Let $G = (V, E)$ be a graph, and V_1, V_2 be non-empty disjoint subsets of V. We define the* density *of the two subsets to be*

$$d(V_1, V_2) = \frac{|e(V_1, V_2)|}{|V_1||V_2|}$$

where $e(V_1, V_2)$ denotes the set of edges between V_1 and V_2.

This is a normalized version of the definition of density we used in the previous lectures.

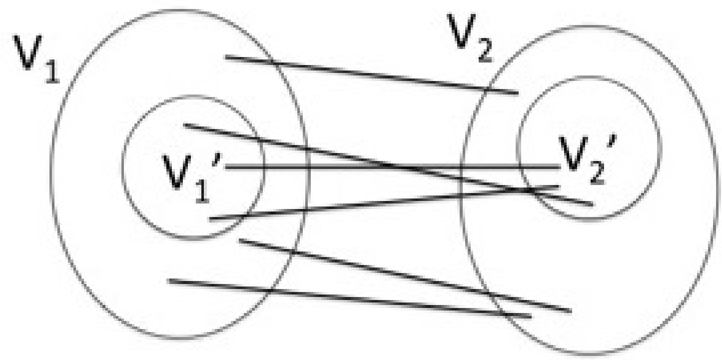

Figure 1: Regularity: "inner" density similar to "outer" density.

Definition 3 (γ-regularity)**.** *A pair of disjoint subsets* (V_1, V_2) *is* γ*-regular if* $\forall V_1' \subseteq V_1, V_2' \subseteq V_2$ *such that* $|V_1'| > \gamma|V_1|$ *and* $|V_2'| > \gamma|V_2|$ *the following holds:* $|d(V_1, V_2) - d(V_1', V_2')| < \gamma$*.*

This definition attempts to capture a "random-like" property of the graph – we expect any two subsets of a random graph to behave in this manner with high probability.

The Regularity Lemma deals with *equipartitions* of a graph, that is, partitions of its vertices into disjoint subsets that differ in size by at most 1. We would like to obtain equipartitions where almost all pairs of subsets are regular. The lemma states that, given any equipartition, one can always partition it further (or *refine* it) to get a new equipartition where almost all pairs of subsets are regular.

Theorem 4 (Szemerédi's Regularity Lemma)**.** $\forall m, \forall \epsilon > 0\ \exists T = T(m, \epsilon)$ *such that if* $G = (V, E)$ *is a graph with more than* T *vertices and* $\mathcal{A}$ *is an equipartition of* V *into* m *sets, then there is an equipartition* $\mathcal{B}$ *of* V *that is a refinement*[1] *of* $\mathcal{A}$ *with* $|\mathcal{B}| = k$ *sets satisfying:*

1. $m \leq k < T$;
2. *at most* $\epsilon\binom{k}{2}$ *pairs of sets in* $\mathcal{B}$ *are not* ϵ*-regular.*

Notice that T in the Regularity Lemma, the upper bound on the number of sets in the "almost regular" equipartition, does not depend on the size of the graph. It only depends on m and ϵ. However, the dependence on ϵ is prohibitively large for any practical applications: it is a tower of height $\Theta(\frac{1}{\epsilon})$. Nevertheless, the Regularity Lemma has a huge theoretical significance.

2.3 Triangles in a Random Tripartite Graph

Consider a random tripartite graph of density at least η. That is, a graph on n nodes constructed by partitioning the nodes into three sets, A, B and C, of $\frac{n}{3}$ nodes each, and for every pair of nodes (u, v) where u and v are in different sets, adding an edge (u, v) with probability at least η (no edges inside each set).

[1] Every set in $\mathcal{B}$ is a subset of a set in $\mathcal{A}$.

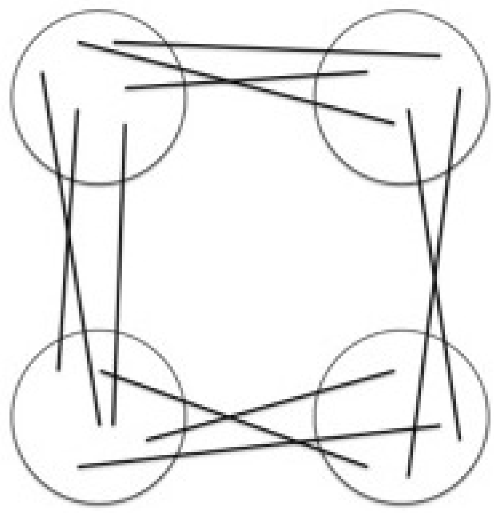

Figure 2: $\mathcal{A}$: A partition of vertices into $m = 4$ sets.

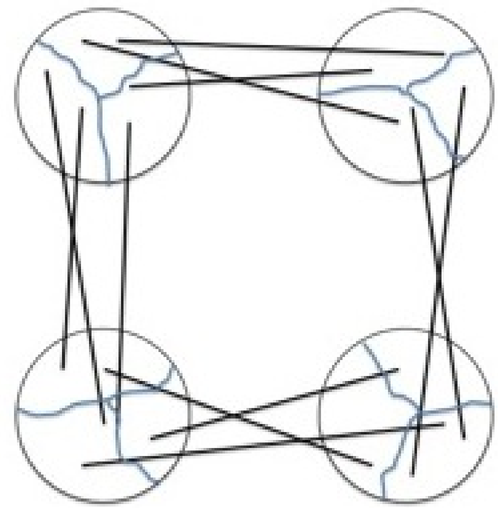

Figure 3: $\mathcal{B}$: A refinement of $\mathcal{A}$ into $k = 3m$ sets.

Question: *How many triangles do we expect to see?*

Let X_{uvw} be the indicator for the event that $\{u, v, w\}$ form a triangle. Then, $E[X_{uvw}] \geq \eta^3$ and by the linearity of expectation

$$E[\sum_{u\in A, v\in B, w\in C} X_{uvw}] = \sum_{u\in A, v\in B, w\in C} [E(X_{uvw}] \geq \left(\frac{n}{3}\right)^3 \eta^3.$$

2.4 Triangles in a Graph with Three Pairwise Regular Pairs

Similarly, we expect to see lots of triangles in a graph with three pairwise regular pairs A, B, C, where each pair of sets has density at least η.

Lemma 5 (Komlos Simonovits, special case for $\triangle$'s [KSSS00])**.** *$\forall \eta > 0$ $\exists \gamma^{\triangle} = \gamma(\eta)$ and $\delta^{\triangle} = \delta(\eta)$ such that if A, B, C are disjoint subsets of V and each pair of them is $\gamma^{\triangle}$–regular with density $\geq \eta$, then G contains $\geq \delta^{\triangle}|A| \cdot |B| \cdot |C|$ distinct triangles with a node from each set. Furthermore, $\gamma^{\triangle} = \frac{\eta}{2}$ and $\delta^{\triangle} = \frac{1}{8}(1-\eta)\eta^3$.*

Proof. Let A^* be the set of nodes in A, each of which has more than $(\eta - \gamma^{\triangle})|B|$ neighbors in B and more than $(\eta - \gamma^{\triangle})|C|$ neighbors in C.

Claim 6. $|A^*| \geq (1 - 2\gamma^{\triangle})|A|$.

Proof of Claim 6: Let A' be the set of nodes in A, each of which has at most $(\eta - \gamma^{\triangle})|B|$ neighbors in B. Then,

$$d(A', B) = \frac{|e(A', B)|}{|A'| \cdot |B|} \leq \frac{|A'|(\eta - \gamma^{\triangle})|B|}{|A'| \cdot |B|} = \eta - \gamma^{\triangle}.$$

By the hypothesis, $d(A, B) \geq \eta$, and we get that

$$|d(A, B) - d(A', B)| \geq \eta - (\eta - \gamma^{\triangle}) = \gamma^{\triangle}.$$

By $\gamma^{\triangle}$–regularity of (A, B), we conclude that $|A'| < \gamma^{\triangle}|A|$ (otherwise $|d(A, B) - d(A', B)| < \gamma^{\triangle}$).

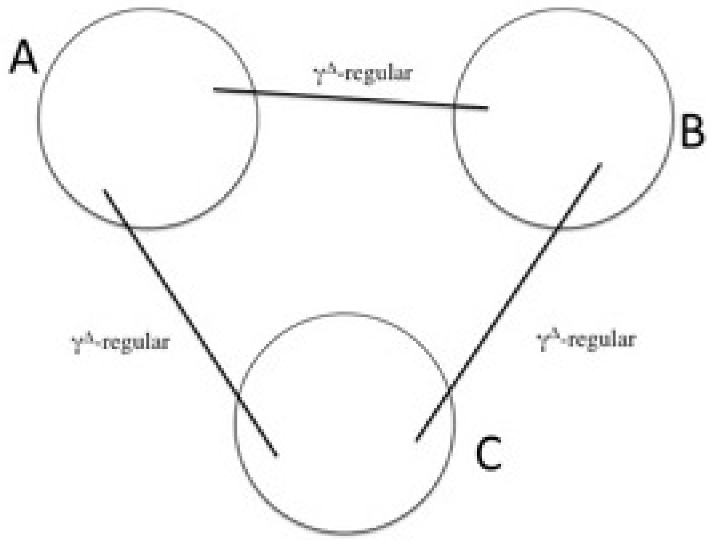

Figure 4: $\gamma^{\triangle}$-regular disjoint subsets of V.

Similarly, let A'' be the set of nodes in A, each of which has at most $(\eta - \gamma^{\triangle})|C|$ neighbors in C, and conclude that $|A''| \leq \gamma^{\triangle}|A|$. Putting the bounds on the sizes of A' and A'' together, we get the claim:

$$|A^*| = |A \backslash \{A' \cup A''\}| \geq (1 - 2\gamma^{\triangle})|A|.$$

□

Back to the proof of the lemma. Let $v \in A^*$ and let B_v denote the set of neighbors of v in B, and C_v, the set of neighbors of v in C. Since we set $\gamma^{\triangle} = \frac{\eta}{2}$,

$$|B_v| > (\eta - \gamma^{\triangle})|B| = \gamma^{\triangle}|B|.$$

Each edge between B_v and C_v contributes a triangle (with vertex v). So, the question is how many edges there are between B_v and C_v.

The pair (B, C) is $\gamma^{\triangle}$-regular with density $d(B, C) \geq \eta$. Therefore, $d(B_v, C_v) > \eta - \gamma^{\triangle}$ and we get:

$$|e(B_v, C_v)| = d(B_v, C_v) \cdot |B_v| \cdot |C_v| > \left(\eta - \gamma^{\triangle}\right)^3 |B| \cdot |C| = \left(\frac{\eta}{2}\right)^3 |B| \cdot |C|.$$

By the claim,

$$|A^*| \geq (1 - 2\gamma^{\triangle})|A| = (1 - \eta)|A|.$$

Setting $\delta^{\triangle} = \frac{1}{8}(1 - \eta)\eta^3$ gives that the number of distinct triangles with a node from each set is

$$\geq \underbrace{(1 - \eta)|A|}_{|A^*|} \underbrace{\left(\frac{\eta}{2}\right)^3 |B| \cdot |C|}_{\#\triangle \forall v \in A^*} = \frac{1}{8}(1 - \eta)\eta^3 |A| \cdot |B| \cdot |C| = \delta^{\triangle}|A| \cdot |B| \cdot |C|.$$

□

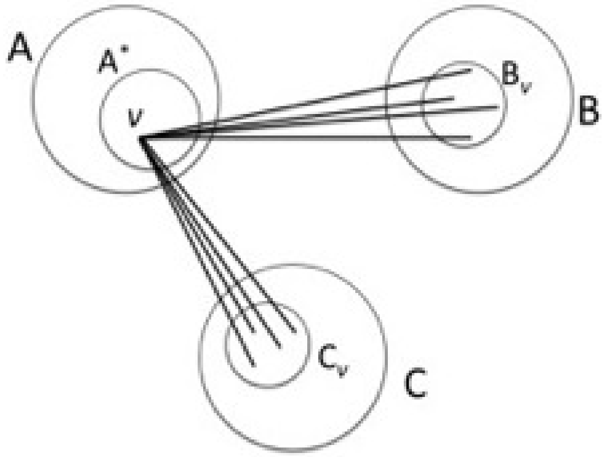

Figure 5: $v \in A^*$ and its neighbors in B and C.

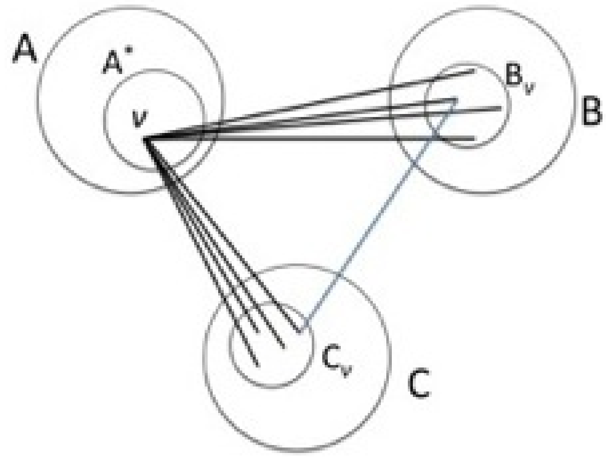

Figure 6: Each edge between B_v and C_v contributes a triangle.

3 Proof of the Triangle Removal Lemma (Theorem 1)

Proof of Theorem 1: Consider the graph G which is ϵ-far from being triangle-free.

We start with an equipartition A of G with $\frac{5}{\epsilon}$ sets.

Set $\epsilon' = \min\{\frac{\epsilon}{5}, \gamma^{\Delta}(\frac{\epsilon}{5})\} = \frac{\epsilon}{10}$. (The $\gamma^{\Delta}(\cdot)$ function is defined in KS Lemma.)

Apply the Regularity Lemma by setting the parameters as follows, $m = \frac{5}{\epsilon}$, $\epsilon = \epsilon'$ (this ϵ represents the parameter ϵ in Regularity Lemma), we have $T = T(\frac{5}{\epsilon}, \epsilon')$. By Regularity Lemma, A can be refined into an equipartition $B = \{V_1, V_2, ..., V_k\}$, such that,

1. $\frac{5}{\epsilon} \leq k \leq T$ (that is, $|V_i| = \frac{n}{k} \in [\frac{n}{T}, \frac{\epsilon n}{5}]$, for $i \in [k]$);
2. at most $\epsilon' \binom{k}{2}$ pairs of $\{V_1, V_2, ..., V_k\}$ are not ϵ'-regular.

Definition 7 (useful edge)**.** *An edge (u, v), where $u \in V_i$, $v \in V_j$, is useful if it satisfies the following three conditions: (1) $i \neq j$; (2) (V_i, V_j) is ϵ'-regular; (3) the density $d(V_i, V_j) \geq \frac{\epsilon}{5}$.*

We show that G does not contain many non-useful edges.

Claim 8. *G has less than $\epsilon\binom{n}{2}$ non-useful edges with respect to equipartition B.*

We complete the proof of the Triangle-Removal Lemma by applying Claim 8. We remove all the non-useful edges in G. Since G is ϵ-far from being triangle-free, there are triangles left in G. Let (u, v, w) be a triangle, where $u \in V_{i_1}$, $v \in V_{i_2}$, $w \in V_{i_3}$.

By KS Lemma, the number of triangles in G is at least $\delta^{\Delta}(\frac{\epsilon}{5})|V_{i_1}||V_{i_2}||V_{i_3}|$, which is at least $\frac{\epsilon}{10}(\frac{n}{T})^3 = \delta \cdot \binom{n}{3}$.

Proof of Claim 8: We bound the number of edges which violate each of the three conditions that a regular edge must satisfy.

1. Let n_1 be the number of edges violating the first condition. Each vertex in G can have at most $\frac{n}{k} - 1$ neighbors in the same partition.

$$n_1 \leq (\frac{n}{k} - 1) \cdot n \leq \frac{(n-1)n}{k} = \frac{2}{k} \cdot \binom{n}{2} \leq \frac{2\epsilon}{5} \cdot \binom{n}{2}$$

.

2. Let n_2 be the number of edges violating the second condition. By Regularity Lemma, there are at most $\epsilon'\binom{k}{2}$ pairs of partitions in B which are not ϵ'-regular. Each of the pairs contributes at most $(\frac{n}{k})^2$ cross edges.

$$n_2 \leq \epsilon'\binom{k}{2} \cdot (\frac{n}{k})^2 \leq \epsilon' \cdot \frac{k(k-1)}{2} \cdot \frac{n(n-1)}{k(k-1)} = \epsilon'\binom{n}{2} \leq \frac{\epsilon}{5} \cdot \binom{n}{2}.$$

3. Let n_3 be the number of edges violating the third condition. For each pair V_i, V_j, $V_i \neq V_j$ and $d(V_i, V_j) < \frac{\epsilon}{5}$, we have $e(V_i, V_j) = d(V_i, V_j)|V_i||V_j|$,

$$n_3 \leq \binom{k}{2} e(V_i, V_j) < \frac{\epsilon}{5} \cdot (\frac{n}{k})^2 \cdot \binom{k}{2} \leq \frac{\epsilon}{5} \cdot \binom{n}{2}.$$

Hence, $n_1 + n_2 + n_3 \leq \frac{4\epsilon}{5}\binom{n}{2} < \epsilon\binom{n}{2}$. This completes the proof of the claim. □ □

References

[AFKS00] Noga Alon, Eldar Fischer, Michael Krivelevich, and Mario Szegedy. Efficient testing of large graphs. *Combinatorica*, 20(4):451–476, 2000.

[AFNS09] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: It's all about regularity. *SIAM J. Comput.*, 39(1):143–167, 2009.

[KSSS00] János Komlós, Ali Shokoufandeh, Miklós Simonovits, and Endre Szemerédi. The regularity lemma and its applications in graph theory. In Gholamreza B. Khosrovshahi, Ali Shokoufandeh, and Mohammad Amin Shokrollahi, editors, *Theoretical Aspects of Computer Science*, volume 2292 of *Lecture Notes in Computer Science*, pages 84–112. Springer, 2000.

[Sze78] Endre Szemerdi. Regular partitions of graphs. In *Problmes Combinatoires et Thorie des Graphes*, volume 260 of *Colloq. Internat. CNRS*, pages 399–401. CNRS, Paris, 1978.