Sublinear Algorithms	February 9, 2012

Lecture 10

Lecturer: Sofya Raskhodnikova | *Scribe(s): Megan Heysham*

1 Introduction

In this lecture we state (but do not prove) the Regularity Lemma. We also present an algorithm to test for triangle-freeness in the dense graph model. 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 exponentations is $1/\epsilon$). The result is of combinatorial interest, but the algorithm is not very practical.

2 Testing △-freeness and the Regularity Lemma

2.1 Testing △-freeness

We look at a special case of a theorem from [AFNS09], the property of △-freeness. The more general proof uses similar techniques.

Input: $G = (V, E)$ undirected graph in the adjacency matrix model

Goal: ϵ-tester for △-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 (△-removal lemma)**.** *$\forall \epsilon$ $\exists \delta = \delta(\epsilon)$ such that if a graph G on n nodes is ϵ-far from △-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 △-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.

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: $|d(V_1, V_2) - d(V_1', V_2')| < \gamma$.*

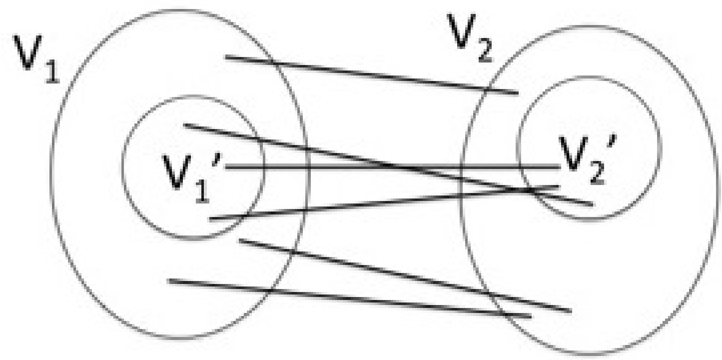

Figure 1: Regularity: "inner" density similar to "outer" density

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 η (no edges inside each set).

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

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

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

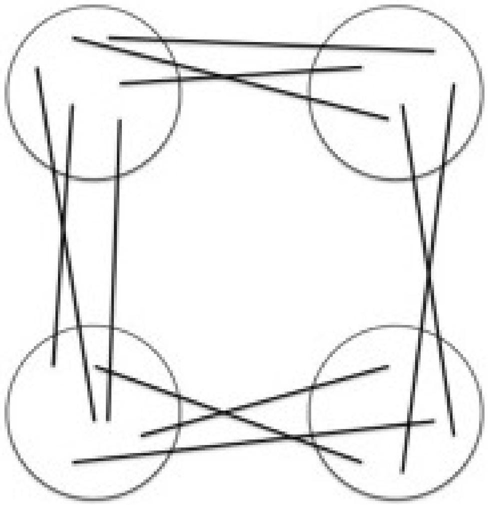

Figure 2: $\mathcal{A}$: A partition of vertices into $m = 4$ sets

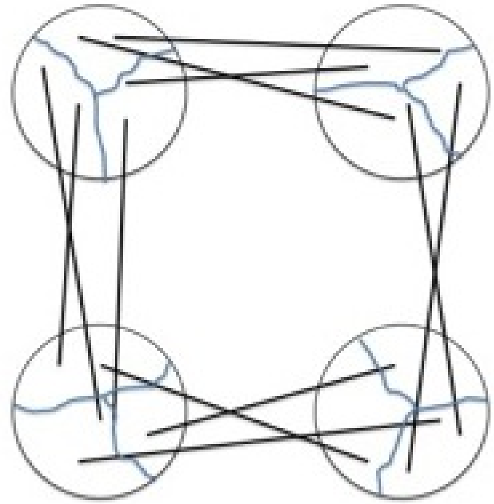

Figure 3: $\mathcal{B}$: A refinement of $\mathcal{A}$ into $k = 3m$ sets

2.4 Triangles in a "regular" tripartite graph

Similarly, we expect to see lots of triangles in a "regular" tripartite graph with density at least η, that is, in a graph where A, B, C are pairwise regular and 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 with at least $(\eta - \gamma^{\triangle})|B|$ neighbors in B and at least $(\eta - \gamma^{\triangle})|C|$ 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 with $< (\eta - \gamma^{\triangle})|B|$ neighbors in B. Then,

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

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

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

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

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

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

□

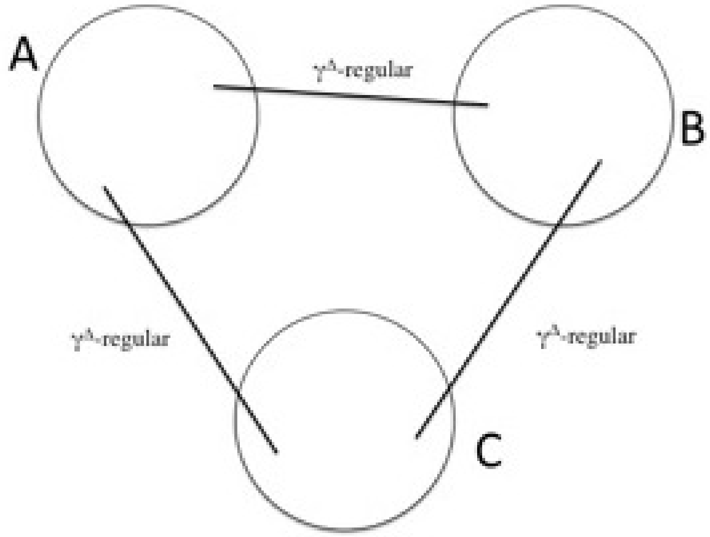

Figure 4: $\gamma^{\triangle}$-regular disjoint subsets of V

Back to the lemma's proof. 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 $\gamma^{\triangle} = \frac{\eta}{2}$,

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

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, $d(B_v, C_v) \geq \eta - \gamma^{\triangle}$ and we get:

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

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

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

□

References

[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.

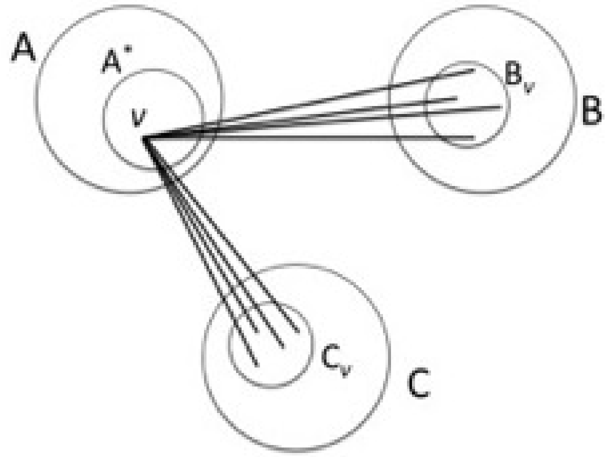

Figure 5: $v \in A^*$ and its neighbors in B and C

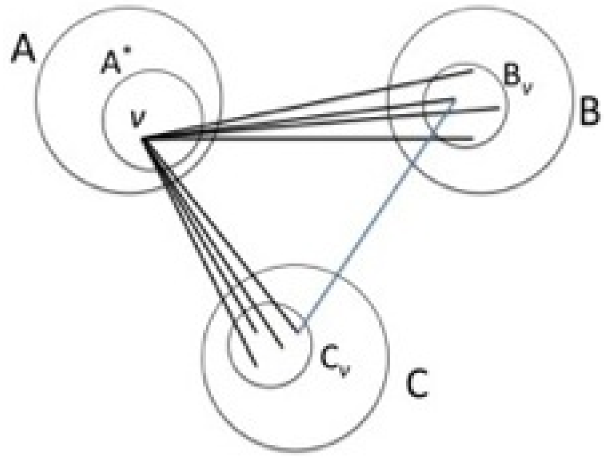

Figure 6: Each edge between B_v and C_v contributes a triangle

[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.