Lecture 5

Lecturer: Elena Grigorescu *Scribe: Abram Magner*

1 Overview

Our goal in the next few lectures shall be to show that triangle-freeness is testable with a constant number of queries. The main tool used in the proof is the triangle removal lemma, which is a consequence of Szemerédi's regularity lemma and the Komlós-Simonovits lemma. We shall state all three of these results and shall prove Komlós-Somonovits.

2 Szemerédi's Regularity Lemma

Roughly speaking, Szemerédi's lemma says that any sufficiently large graph can be partitioned into a small number of components such that any pair of components looks like a random bipartite graph.

While our main interest in it is in proving the triangle removal lemma, it is of independent interest, because it is useful in many combinatorics applications (e.g., Szemerédi's theorem, which settles a conjecture by Erdős and Turán on arithmetic progressions) and because it can be generalized to objects other than graphs (see, for example, Green's theorem on abelian groups) [3]. Furthermore, it forms the basis of a combinatorial characterization of the set of graph properties that are testable with a query complexity constant in the size of the graph [1].

To state it precisely, we'll need some definitions.

Given vertex sets X and Y, let $e(X, Y)$ denote the number of edges with one end point in X and the other in Y. Given disjoint vertex sets V_1 and V_2, define the density of the pair to be

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

A complete bipartite graph on vertex sets V_1 and V_2 has $|V_1||V_2|$ edges, so the density is a number in the interval $[0, 1]$.

Definition 1 (ϵ-regular) *A bipartite graph $G = (V_1 \cup V_2, E)$ is called* ϵ-regular *if, for all $X \subseteq V_1$ and $Y \subseteq V_2$ satisfying $|X| \geq \epsilon|V_1|$ and $|Y| \geq \epsilon|V_2|$, $|d(X, Y) - d(V_1, V_2)| \leq \epsilon$.*

In other words, G is ϵ-regular if any pair of sufficiently large subsets, one from V_1 and the other from V_2, has density similar to that of V_1 and V_2. For this reason, ϵ-regularity is called a *homogeneity condition* or an *expansion property.*

Definition 2 (Equipartition) *Given a graph $G = (V, E)$, a collection $B = \{V_1, V_2, \ldots, V_k\}$ of vertex sets is called an* equipartition *if B is a partition of V and, for all i and j, $||V_i| - |V_j|| \leq 1$.*

That is, an equipartition of a graph is a paritition whose components all have almost the same cardinality.

Finally, given disjoint vertex sets S_1 and S_2, denote by $G[S_1, S_2]$ the bipartite graph whose edges are precisely those with one endpoint in each of S_1, S_2.

We can now state the lemma.

Lemma 1 (Szemerédi's Regularity Lemma) *For all $\epsilon > 0$, there exists $M = M(\epsilon)$ such that any graph G on n vertices, $n \geq M$, can be equipartitioned into k parts $V_1, V_2, \ldots, V_k$, with $\frac{5}{\epsilon} < k < M(\epsilon)$, such that at least $(1-\epsilon)\binom{k}{2}$ graphs $G[V_i, V_j]$ are ϵ-regular.*

How large is $M(\epsilon)$? For the lemma presented here, it is ${}^{(1/\epsilon)}2$ (i.e., a tower of 2 of height $\frac{1}{\epsilon}$). Frieze and Kannan give $M(\epsilon) = 2^{c/\epsilon}$ for a modified notion of regularity [2].

The idea of the proof of the lemma is to define a norm for a vertex partition such that, if we're given a partition that isn't sufficiently regular, we can construct a refined partition whose norm differs from that of the original by at least a constant ϵ. By giving an upper bound on the norm of any partition, we can use this to show that the number of refinements must be finite and that the resulting partition has the claimed regularity property.

3 Triangle Removal Lemma

Our interest in the regularity lemma is, ultimately, a result of the fact that it is useful in showing that triangle-freeness is testable (in the dense graph model) with a number of queries constant in the size of the input graph. The problem can be stated as follows.

We want an algorithm A such that, given a graph G and $\epsilon > 0$, A

1. makes $q(\epsilon)$ queries (i.e., the query complexity is constant with respect to the size of the graph),
2. accepts if G is triangle-free,
3. and rejects with probability at least $\frac{2}{3}$ if G is ϵ-far from triangle-free.

Here, G is ϵ-far from triangle-free if and only if at least $\epsilon\binom{n}{2}$ edges must be removed from G in order to render it triangle-free.

How do we test for triangle-freeness? It turns out that the following simple idea works: without replacement, pick 3 vertices uniformly at random. If they form a triangle, then reject. After repeating this test some number (call it s, the value of which we will determine in the next lecture) of times, if no triangles were detected, accept.

Clearly, if G is triangle-free, the above test cannot reject, so it is one-sided. If G is ϵ-far, then the probability that the simple test rejects is

$$\Pr[\text{reject}] = \frac{\Delta(G)}{\binom{n}{3}} \tag{1}$$

Here, $\Delta(G)$ denotes the number of triangles in G. To complete the analysis, we need an estimate for $\Delta(G)$. Since G is ϵ-far, we know that it contains at least $\Omega(\epsilon n^2)$ triangles, but the triangle removal lemma tells us more.

Lemma 2 (Triangle Removal Lemma) *For all $\epsilon > 0$, there exists $\delta(\epsilon)$ such that if G is a graph on n vertices that is ϵ-far from triangle-free, then G has at least $\delta \cdot \binom{n}{3}$ distinct triangles.*

The proof of the lemma, which will be given in the next lecture, relies on Szemerédi's lemma and on the Komlós-Simonovits lemma, discussed next.

4 Komlós-Simonovits Lemma

Consider a random tripartite graph $G = V_1 \cup V_2 \cup V_3$, where each $|V_i| = \frac{n}{3}$, such that any edge appears with probability η, independently of all other edges. The expected number of triangles can be computed as follows. After fixing some ordering of the vertices,

$$\mathbb{E}[\Delta(G)] = \mathbb{E}\left[\sum_{u<v<w} \mathbb{I}[uvw \text{ is a triangle }]\right] = \sum_{u<v<w} \Pr[uvw \text{ is a triangle }] \tag{2}$$

$$= \eta^3 \binom{n}{3}^3 = \Theta(n^3), \tag{3}$$

where each sum is over all ordered triples of vertices. We expect a graph G partitioned into three subsets such that each pair of subsets is sufficiently regular and dense to behave in a "random-like" fashion, so we might expect the number of triangles in G to be similar to the expected number just computed. The Komlós-Simonovits lemma formalizes this.

Lemma 3 (Komlós-Simonovits) *If $G = A \cup B \cup C$ (with A, B, and C pairwise disjoint) is such that $(A, B), (A, C)$, and (B, C) are $\frac{\eta}{2}$-regular and $d(A, B), d(A, C), d(B, C) > \eta$, then G has at least $\delta|A||B||C|$ triangles, where $\delta = \frac{1}{8}(1-\eta)\eta^3$.*

Proof First, for all vertices x, let B_x and C_x denote the neighborhoods of x in B and C, respectively (usually denoted by $\Gamma_x(B)$ and $\Gamma_x(C)$). Now, consider $a \in A$ such that $|B_a| > \frac{\eta}{2}|B|$ and $|C_a| > \frac{\eta}{2}|C|$. Intuitively, a is chosen so that it has many neighbors in B and C. We will prove later that there is a sufficiently large number of such a.

We claim that $e(B_a, C_a) > \left(\frac{\eta}{2}\right)^3 |B||C|$. Since (B, C) is a $\frac{\eta}{2}$-regular pair and $|B_a| > \frac{\eta}{2}|B|$ and $|C_a| > \frac{\eta}{2}|C|$, we can conclude that $|d(B_a, C_a) - d(B, C)| < \frac{\eta}{2}$. Since, by hypothesis, $d(B, C) > \eta$, this implies that $d(B_a, C_a) > \frac{\eta}{2}$, so that $e(B_a, C_a) > \frac{\eta}{2}|B_a||C_a| > \left(\frac{\eta}{2}\right)^3 |B||C|$, as desired.

Now, we need to count the number of a satisfying the properties that we defined. Let $A' = \{a \in A : |B_a| > \frac{\eta}{2}|B| \text{ and } |C_a| > \frac{\eta}{2}|C|\}$. The claim is that $|A'| > (1-\eta)|A|$. Let A_B be the set of $a \in A$ such that $|B_a| \le \frac{\eta}{2}|B|$, and define A_C analogously. To prove the claim, we will show that both of these sets are small, so that their union (i.e., the complement of A' in A) cannot be too large. We start with

$$d(A_B, B) = \frac{e(A_B, B)}{|A_B||B|} \le \frac{|A_B|\frac{\eta}{2}|B|}{|A_B||B|} = \frac{\eta}{2}. \tag{4}$$

The inequality follows from the definition of A_B: each element $a \in A_B$ has at most $\frac{\eta}{2}|B|$ connections to vertices in B, so that the total number of edges between A_B and B is at most $|A_B|\frac{\eta}{2}|B|$.

This implies that $|d(A, B) - d(A_B, B)| \ge d(A, B) - d(A_B, B) > \eta - \frac{\eta}{2} = \frac{\eta}{2}$. Since (A, B) is $\frac{\eta}{2}$-regular, it must be that $|A_B| < \frac{\eta}{2}|A|$. Exactly the same argument works for A_C, so that $|A_C| < \frac{\eta}{2}|A|$.

The above bounds imply that $|A - A'| = |A_B \cup A_C| \le |A_B| + |A_C| < 2 \cdot \frac{\eta}{2}|A| = \eta|A|$, which further implies that $|A'| = |A| - |A - A'| > (1-\eta)|A|$, as claimed.

Bringing everything together, if we let Δ denote the number of triangles in G,

$$\Delta > |A'| \max_{a \in A} e(B_a, C_a) \ge (1-\eta)|A| \left(\frac{\eta}{2}\right)^3 |B||C| = \frac{1}{8}(1-\eta)\eta^3|A||B||C|, \tag{5}$$

as desired. ■

References

[1] Noga Alon, Eldar Fischer, Ilan Newman, and Asaf Shapira. A combinatorial characterization of the testable graph properties: its all about regularity. In *Proc. of STOC 2006*, pages 251–260, 2006.

[2] A. Frieze and R. Kannan. The regularity lemma and approximation schemes for dense problems. In *Foundations of Computer Science, 1996. Proceedings., 37th Annual Symposium on*, pages 12 –20, oct 1996.

[3] B. Green. A szemeredi-type regularity lemma in abelian groups, with applications. *Geometric And Functional Analysis*, 15:340–376, 2005. 10.1007/s00039-005-0509-8.