Lecture 6. Key lemmas

Hong Liu

2nd April 2021

In the last lecture we have seen how we can reduce a problem on a large graph G to one on its reduced graph R. In this section, we shall formally define this reduced graph, state and prove some key lemmas related to regularity lemma, in particular, the embedding lemma, the counting lemma and the triangle removal lemma.

Roughly speaking, the embedding lemma says that we can embed any (appropriate) bounded degree graphs (up to linear-size); and the counting lemma says for any fixed (small) graph H, we can count accurately the number of copies of H in G. We remark that there is a stronger version of embedding lemma, the blow up lemma, due to Komlós, Sárközy and Szemerédi, which we will not cover for now. The blow up lemma states that we can embed any (appropriate) spanning bounded degree graphs.

1 Reduced graph

Here is the formal definition of reduced graphs obtained from applying regularity lemma to a large graph.

Definition 1.1 (Reduced graph)**.** Given an ε-regular partition $V(G) = V_0 \cup V_1 \cup \cdots \cup V_r$ of G, and $\delta > 0$, the *reduced/cluster graph* $R = R(\varepsilon, \delta)$ of G is defined as follows:

- $V(R) = [r]$;
- $ij \in E(R)$ if and only if (V_i, V_j) is ε-regular with density at least δ.

It is often helpful to think of a reduced graph as a weighted graph, assigning weight

$$d_{ij} := d(V_i, V_j)$$

to the edge ij in R, and define the weighted degree of vertex $i \in V(R)$ to be

$$d_R(i) = \sum_{j \in N_R(i)} d_{ij}.$$

We will specify it when we treat R as a weighted graph.

Exercise 1.2. Normalised minimum degree is inherited by the reduced graph $R = R(\varepsilon, \delta)$, i.e.[1]

$$\frac{\delta(R)+1}{r} \geq \frac{\delta(G)}{n} - \delta - \varepsilon \left(= \frac{\delta(G)}{n} - o(1) \right).$$

Exercise 1.3. Bound the edge-density of G by that of R's, i.e.

$$\frac{e(G)}{\binom{|G|}{2}} \leq \frac{e(R)}{\binom{|R|}{2}} + o(1).$$

As we shall soon see in the counting lemma, the notion of reduced graph $R(\varepsilon, \delta)$ captures essentially the whole (asymptotic) information of G on subgraphs densities.

[1] For this, we use the degree version of the regularity lemma. That is, we can assume that each part V_i, $i \in [r]$, is in at most εr irregular pairs.

2 Embedding lemma

Given a graph F, denote by $F(s)$ the *blow-up* of F obtained from replacing each vertex $u \in V(F)$ by an independent set I_u of size s and make (I_u, I_v) complete bipartite in $F(s)$ if and only if $uv \in E(F)$. Observe that

the blow-up $F(s)$ contains H as a subgraph $\iff$ F contains a homomorphic copy of H.

We now present the embedding lemma, which is a formal statement of property that if a graph is H-free, then its reduced graph does not contain any homomorphic image of H. One thing to notice here is that we can embed appropriate bounded degree graphs of order up to linear size $\Omega(n)$ (so think of d, Δ below as constants and $|G| = \Theta(\ell)$, $|H| = \Theta(s)$ and $s = \Omega(\ell)$). Later we will embed a graph of order linear in the order of the host graph in Chvátal-Rödl-Szemerédi-Trotter theorem about Ramsey number of bounded degree graphs.

Lemma 2.1 (Embedding lemma)**.** *For any $d \in (0, 1]$, $\Delta \geq 1$, there exists $\varepsilon_0 > 0$ such that the following holds for any $\varepsilon \leq \varepsilon_0$. Let G and H be graphs with $\Delta(H) \leq \Delta$, $s \in \mathbb{N}$ and $R = R(\varepsilon, d)$ be a reduced graph of G. Suppose the corresponding regular partition of G has each of its part of size $\ell \geq 2s/d^{\Delta}$. Then*

$$H \subseteq R(s) \quad \Rightarrow \quad H \subseteq G.$$

Sketch of proof. Given d, Δ, choose $\varepsilon_0 < d$, such that

$$(d - \varepsilon_0)^{\Delta} - \varepsilon_0 \Delta \geq \frac{1}{2} d^{\Delta} \geq \varepsilon_0.$$

Let $\varphi : V(H) \to V(R)$ be a homomorphism (exists as $H \subseteq R(s)$). Order vertices in H as $u_1, \ldots, u_h$. Initially, set $Y_j = V_j$, for $j \in [r]$. Embed vertices $u_1, \ldots, u_{i-1}$ one by one, and update the sets of eligible vertices $Y_{\varphi(u_j)} \subseteq V_{\varphi(u_j)}$ for each u_j, $j \geq i$ and $u_{i-1}u_j \in E(H)$, to embed by intersecting it with $N(u_{i-1})$, maintaining always $|Y_{\varphi(u_j)}| \geq \varepsilon |V_{\varphi(u_j)}|$. When embedding u_i in $Y_{\varphi(u_i)} \subseteq V_{\varphi(u_i)}$, note that for each $j > i$ with $u_i u_j \in E(H)$, in $V_{\varphi(u_i)}$, all but $\varepsilon |V_{\varphi(u_i)}|$ vertices u, by Lemma 2.5 from last lecture, satisfy $d(u, Y_{\varphi(u_j)}) \geq (d - \varepsilon)|Y_{\varphi(u_j)}|$. Since

$$|V_i|(d - \varepsilon)^{\Delta} - \varepsilon \Delta |V_i| \geq \max\{s, \varepsilon |V_i|\},$$

we never get stuck. □

Exercise 2.2. Make the proof of the upper bound of Erdős-Simonovits-Stone theorem rigorous.

3 Counting lemma

The counting lemma below states that a (weighted) reduced graph preserves subgraph densities. Notice that the count given below is what we would expect if graphs between pairs (V_i, V_j) are completely random.

Lemma 3.1 (Counting lemma)**.** *Given H, $V_1, ..., V_h$ with $h = |H|$ and $|V_i| = n$, all pairs (V_i, V_j) are ε-regular and $d(V_i, V_j) = d_{ij} \gg \varepsilon$. Then the number of canonical copies[2] of H in $V_1, ..., V_h$ is at least*

$$\prod_{ij \in E(H)} (d_{ij} - C\varepsilon) n^h,$$

where $C = C(H)$ is a constant depending only on H.

We skip the proof for the counting lemma, instead leaving the baby case of triangle counting as exercise.

[2]By canonical copy, we mean a copy of H with exactly one vertex in each V_i.

Exercise 3.2. Prove counting lemma for the special case $H = K_3$.

One can also count induced copies of a fixed size graph.

Exercise 3.3. Formulate and prove a counting lemma for induced C_4.

4 Ruzsa-Szemerédi triangle removal lemma

In this section, we will present, yet, another important consequence of the regularity lemma, the triangle removal lemma, due to Ruzsa and Szemerédi, which states that an almost triangle-free graph ($o(n^3)$ triangles) can be made genuinely triangle-free by removing a negligible amount of edges ($o(n^2)$ edges).

Lemma 4.1 (Ruzsa-Szemerédi triangle removal lemma 1976)**.** *Given $c > 0$, there exists $a = a(c) > 0$, such that for sufficiently large n the following holds. Let G be an n-vertex graph. If G has at most an^3 triangles, then it can be made triangle-free by removing at most cn^2 edges.*

The contrapositive says if one cannot make a graph triangle-free by removing few edges (i.e. there are many edge-disjoint triangles), then the graph contains lots (positive proportion) of triangles. The removal lemma has many applications, e.g. $(6, 3)$-theorem and Roth's theorem on size of sets without 3-term arithmetic progressions.

4.1 Cleaning the graph G

Before proving the removal lemma, It is convenient to define the subgraph G_R of G corresponding to a reduced graph $R = R(\varepsilon, \delta)$, obtained by keeping only edges between (regular and dense) pairs (V_i, V_j) for which $ij \in E(R)$. We can obtain the subgraph G_R via the following standard cleaning process, showing that only a negligible amount of edges are deleted:

$$e(G_R) = e(G) - o(n^2).$$

- Remove inner edges, i.e. edges in V_i, $i \in [r]$. By choosing $m \geq 1/\varepsilon$ when applying the regularity lemma to obtain the regular partition corresponding to $R(\varepsilon, \delta)$, we can guarantee the number of parts satisfies $r \geq m \geq 1/\varepsilon$. Then the number of inner edges is at most

$$\binom{n/r}{2} \cdot r \leq \frac{n^2}{2r} \leq \frac{n^2}{2m} = \frac{1}{2}\varepsilon n^2.$$

- Remove edges between irregular pairs. As there are at most εr^2 irregular pairs, the number of edges of this kind is at most

$$\varepsilon r^2 \cdot (n/r)^2 = \varepsilon n^2.$$

- Remove edges between sparse pairs with density at most δ, i.e. (V_i, V_j) with $ij \notin E(R)$. The number of such edges is at most

$$\delta \left(\frac{n}{r}\right)^2 \binom{r}{2} \leq \frac{1}{2}\delta n^2.$$

Thus, in forming G_R, we delete in total at most

$$\frac{1}{2}(3\varepsilon + \delta)n^2 = O(\varepsilon + \delta)n^2$$

edges, which is negligible as we usually choose ε, δ sufficiently small.

The edges deleted in the cleaning process above is exactly the non-essential information we discard when forming the reduced graph. Edges in G_R all lie in regular and dense pairs and so we can employ e.g. the counting lemma, which is how we shall prove the triangle removal lemma.

4.2 Proof of triangle removal lemma

Suppose the statement is not true. That is, there is some $c > 0$ such that for any a there exists a counterexample G, i.e. G has at most an^3 triangles, but the removal of any cn^2 edges does not make it triangle-free.

Apply Szemerédi's regularity lemma with $\varepsilon \ll c$ and $m = 1/\varepsilon$ to G to get an ε-regular partition $V(G) = V_1 \cup ... \cup V_r$, where $M \geq r \geq m$ and $||V_i| - |V_j|| \leq 1$, for $1 \leq i, j \leq r$. Let $R = R(\varepsilon, c/4)$ be the reduced graph, and $G_R \subseteq G$ be the cleaned subgraph, as in Section 4.1. Then the number of edges deleted is at most $cn^2/2$.

By the choice of G, there are still triangles in G_R. Since edges inside V_i and edges between sparse and irregular pairs have been deleted in forming G_R. Each triangle in G_R must have its three vertices lying in three distinct clusters, say X, Y, Z, that are pairwise regular with density larger than $c/4$. We can then apply the counting lemma, Lemma 3.1, to the tripartite graph $G_R[X, Y, Z]$ to see that there are at least

$$\left(\frac{c}{4} - O(\varepsilon)\right)^3 \cdot \left(\frac{n}{r}\right)^3 \geq \left(\frac{c}{8r}\right)^3 n^3 \geq \left(\frac{c}{8M}\right)^3 n^3$$

triangles. Note that $M = M(\varepsilon, m)$ depends in fact only on c. Then choosing $a = a(c) < \left(\frac{c}{8M}\right)^3$, we get that G has more than an^3 triangles, a contradiction.

Remark 4.2. Let us write a streamlined proof for triangle removal lemma without all the calculations. Let G be an almost triangle-free graph. Then its reduced graph R must be triangle-free, as otherwise, by the counting lemma, G would contain too many triangles. Thus, G can be made triangle-free by removing few edges not corresponding to R.