Chapter 4

本讲我们来讨论 Szemerédi Regularity lemma(以后简称 Regularity lemma). 作为图论中最重要的结果之一, Regularity lemma 能够很好的给出关于边比较多的图的一个形状的刻画. 事实上, Regularity lemma 指出, 如果图 G 的边比较多, 我们就能将 G 划分为若干个部分, 任意两个部分之间的连边具有某种" 随机性".

我们首先来给出一些定义并陈述这个定理.

Definition. 给定图 $G(V, E)$. 对于任意两个 V 的不同子集 A, B, 定义 A, B 之间的边密度为

$$d(A, B) = \frac{|e(A, B)|}{|A||B|},$$

其中 $e(A, B)$ 为图 G 中两端点分别在 A 与 B 之间的边数.

Definition. 在图 $G(V, E)$ 中, V 的两个不交子集 (A, B) 被称为是 ϵ-*regular* 的, 如果对于任意满足 $|A'| \geqslant \epsilon|A|$ 的 A 的子集 A', $|B'| \geqslant \epsilon|B|$ 的 B 的子集 B', 均有

$$|d(A', B') - d(A, B)| \leqslant \epsilon.$$

我们可以从定义中看到, ϵ-*regular* 的定义即 A, B 之间的边密度和它们子部分之间的边密度差距很小, 即 A, B 之间的结构具有某种正规性.

接下来我们陈述 Regularity lemma.

Theorem. 对每个实数 $0 < \epsilon < 0.1$, 存在正整数 $M = M(\epsilon)$, 满足: 对任意图 $G(V, E)$, 只要 $|V| \geq M(\epsilon)$, 就存在 V 的分划 $V = V_0 \cup V_1 \cup V_2 \cup \cdots \cup V_t$, 满足:

- $V_1, V_2, \cdots, V_t$ 的元素个数一样;
- $|V_0| < \epsilon n$;
- $\frac{1}{\epsilon} < t < M$;
- 至多 ϵt^2 对 (V_i, V_j) 不是 ϵ-*regular* 的.

我们可以看到, Regularity lemma 指出, 我们可以把 G 的顶点集去掉一个很小的部分后, 将剩下的顶点均匀的分成若干个两两之间几乎都是 ϵ-*regular* 的部分. 有时候我们简称这种划分为 ϵ-*regular* 的划分.

我们首先简单给出一个关于 Regularity lemma 的证明的概要. 首先, 我们取一个同时满足条件 123 的划分, 通过考察一个能够刻画分划正规性的量来进行递降, 逐渐地对原分划进行加细, 并一直保证分划是好的, 最终得到一个满足条件 4 的分划. 我们的操作过程能同时保证条件 1,2,3 成立.

我们来首先对这样的量进行刻画. 下面的定义是关键的:

Definition. 给定 n 个顶点的图 G. 对于任意两个 V 的不同子集 U, W, 定义

$$q(U, W) = \frac{|U||W|}{n^2} d^2(U, W);$$

对于 U 的一个划分 $P_U = \{U_1, U_2, \cdots, U_k\}$, W 的一个划分 $P_W = \{W_1, W_2, \cdots, W_l\}$, 定义

$$q(P_U, P_W) = \sum_{i=1}^{k} \sum_{j=1}^{l} q(U_i, V_j).$$

特殊地, 对于 $V(G)$ 的一个划分 $\mathcal{P}$, 定义划分 $\mathcal{P}$ 的能量 $q(\mathcal{P}) = q(\mathcal{P}, \mathcal{P})$.

我们来解释一下这个定义. 我们随机取 $x \in U, y \in W$, 并考察随机变量 $Z = q(U_i, V_j)$, 其中 $x \in U_i, y \in V_j$. 那么,

$$\mathbb{E}(Z) = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{|U_i|}{|U|} \frac{|W_j|}{|W|} d(U_i, W_j) = \frac{e(U, W)}{|U||W|} = d(U, W),$$
$$\mathbb{E}(Z^2) = \sum_{i=1}^{k} \sum_{j=1}^{l} \frac{|U_i|}{|U|} \frac{|W_j|}{|W|} d(U_i, W_j)^2 = \frac{n^2}{|U||W|} q(P_U, P_W).$$

由于 $\mathbb{E}(Z^2) \geq \mathbb{E}(Z)^2$, 我们就有 $q(P_U, P_W) \geq q(U, W)$. 这也立刻说明, 对于一个分划 $\mathcal{P}$ 和它的加细 $\mathcal{P}'$, 有 $q(\mathcal{P}) \geq q(\mathcal{P}')$, 也就是说, 划分的能量是一个在加细中不减的量.

我们接下来研究 $q(\mathcal{P})$ 怎么刻画正规性. 若 U, W 不 ϵ-*regular*, 设 U, W 的两个子集 U', W' 满足

$$|U'| \geq \epsilon |U|, |W'| \geq \epsilon |V|, |d(U, W) - d(U', W')| \geq \epsilon.$$

那么, 我们考虑 U 的划分 $P_U = \{U', U \setminus U'\}$, W 的划分 $P_W = \{W', W \setminus W'\}$, 则

$$\begin{aligned}
q(P_U, P_W) - q(U, W) &= \frac{|U||W|}{n^2} (\mathbb{E}(Z^2) - \mathbb{E}(Z)^2) \\
&= \frac{|U||W|}{n^2} (\mathbb{E}(Z - \mathbb{E}(Z))^2) \\
&\geq \frac{|U||W|}{n^2} \frac{|U'|}{|U|} \frac{|W'|}{|W|} (d(U', W') - d(U, W))^2 \\
&\geq \epsilon^4 \frac{|U||W|}{n^2}.
\end{aligned}$$

也就是说, 对于不正规的部分, 能量在划分中将会逐渐增大. 然而, 设 $\mathcal{P} = \{V_1, V_2, \cdots, V_m\}$, 则

$$q(\mathcal{P}) = \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{|V_i||V_j|}{n^2} d^2(V_i, V_j) \leq \sum_{i=1}^{m} \sum_{j=1}^{m} \frac{|V_i||V_j|}{n^2} = 1,$$

因此, 能量不可能一直增长, 即我们总可以找到满足条件的分划. 具体来说, 我们需要下面的引理:

Lemma. 给定 V 的分划 $\mathcal{P}=\{V_0,V_1,V_2,\cdots,V_k\}$, 其中 $k\epsilon^6>1$, 且 $|V_1|=\cdots=|V_k|$. 若存在 ϵk^2 对 (V_i,V_j) 不是 *ϵ-regular* 的, 则存在好分划 $\mathcal{P}'''=\{V_0''',V_1''',V_2''',\cdots,V_{k_3}'''\}$, 满足: $|V_1'''|=\cdots=|V_{k_3}'''|$, $k\le k_3\le k4^k$, $|V_0'''|<|V_0|+\frac{n}{2^k}$, 且 $q(\mathcal{P}''')>q(\mathcal{P})+\frac{1}{4}\epsilon^5$.

显然, 我们只要证明了上述引理, 就可以从任意一个满足 $k\epsilon^6>1$ 的分划开始 ($|V|\ge M(\epsilon)$ 保证了这样的分划存在), 逐步操作, 就得到了一个满足 Regularity lemma 的分划. 注意, $k\epsilon^6>1$ 的条件容易说明 Regularity lemma 中的条件 2 始终满足; 而操作次数的有限性和开始时 $k\epsilon^6>1$ 也保证了条件 3.

我们分三步来进行. 不妨设 $|V_1|=m$. 对于每一对不是 *ϵ-regular* 的 V_i,V_j, 我们总是可以选取 $V_i'\subseteq V_i$, $V_j'\subseteq V_j$, 满足

$$|V_i'|\ge\epsilon|U|,|V_j'|\ge\epsilon|V_j|,|d(V_i,V_j)-d(V_i',V_j')|\ge\epsilon.$$

我们将 V_i 分为 V_i' 和 $V_i\setminus V_i'$, V_j 分为 V_j' 和 $V_j\setminus V_j'$, 并记这个 V_i 的划分为 $P_{i,j}$, V_j 的划分为 $P_{j,i}$. 我们同时对所有的 V_i,V_j 执行上述操作 (注意, 一个 V_i 可能被同时执行多次, 执行 t 次就划分为 2^t 个子集合, 这里划分多次按维恩图的方式划分), 得到一个 V 的分划 $\mathcal{P}'$. 我们来比较 $q(\mathcal{P}')$ 与 $q(\mathcal{P})$. 注意到, 对于任意一对若 V_i,V_j 不是 *ϵ-regular* 的, 设 V_i 最终被划分为 P_i, V_j 最终被划分为 P_j, 则

$$q(P_i,P_j)\ge q(P_{i,j},P_{j,i})\ge q(V_i,V_j)+\frac{|V_i||V_j|}{n^2}=q(V_i,V_j)+\epsilon^4\frac{m^2}{n^2},$$

注意, P_i, P_j 均表示分划. 其中第一个不等式是因为 P_i 是 $P_{i,j}$ 的加细, P_j 是 $P_{j,i}$ 的加细. 若 V_i,V_j 是 *ϵ-regular* 的, 也有 $q(P_i,P_j)\ge q(V_i,V_j)$. 将所有这些不等式求和, 我们有

$$q(\mathcal{P}')=\sum q(P_i,P_j)\ge\sum q(V_i,V_j)+\epsilon^4t^2\frac{m^2}{n^2}\ge q(\mathcal{P})+\frac{1}{2}\epsilon^5,$$

其中最后一个不等式是由于 $mt=(1-\epsilon)n>\frac{\sqrt{2}}{2}n$.

设 $\mathcal{P}'=\{V_0',V_1',V_2',\cdots,V_{k_1}'\}$. 则 $k_1\le k2^k$. 我们接下来对 $\mathcal{P}'$ 再进行加细: 将 $V_1',V_2',\cdots,V_{k_1}'$ 中元素个数大于 $\frac{n}{2^kk_1}$ 的进行加细, 将其按 $[\frac{n}{2^kk_1}]$ 个一组分 (可能会剩下若干个元素个数不超过 $[\frac{n}{2^kk_1}]$ 的), 得到一个新的划分 $\mathcal{P}''=\{V_0'',V_1'',V_2',\cdots,V_{k_2}''\}$. 注意这里 V_i'' 由许多元素个数为 $[\frac{n}{2^kk_1}]$ 的集合, 和不超过 k_1 个元素个数小于 $[\frac{n}{2^kk_1}]$ 的集合构成. 此时, $k_2\le 2^kk_1\le k4^k$, $q(P'')\ge q(P')$.

我们接下来把 $V_1'',V_2'',\cdots,V_{k_2}''$ 中元素个数不等于 $\frac{n}{k_1^2}$ 的全部并入 V_0 得到分划 $\mathcal{P}'''=\{V_0''',V_1''',V_2''',\cdots,V_{k_3}'''\}$. 那么, $|V_1'''|=\cdots=|V_{k_3}'''|$; $k_3\le k_2\le k^24^k$; $|V_0'''|\le|V_0|+k_1\frac{n}{2^kk_1}\le|V_0|+\frac{n}{2^k}$. 于是我们只需验证 $q(P''')+\frac{1}{4}\epsilon^5>q(P'')$.

我们将 $q(P'')$ 中元素个数不等于 $\frac{n}{2^kk_1}$ 的集合全部加细为单元集得到分划 $q(P'''')$, 此时这样的单元集个数不超过 $\frac{n}{2^k}$, 故不超过 $\frac{1}{4}\epsilon^5n$. 下证 $q(P''')+\frac{1}{4}\epsilon^5>q(P'''')$. 注意到 $q(P'''')$ 是从 $q(P''')$ 中分出若干个单元集得到的加细, 故我们只需证明, 从任意划分 $\mathcal{F}$ 中的某个集合中分出一个单元集, 能量至多增加 $\frac{1}{n}$. 设 $\mathcal{F}=\{F_0,F_1,F_2,\cdots,F_r\}$. 从 F_0 中分出一个元素 a 后, 能量的变化 S 为

$$S=\sum_{i=1}^{r}(q(F_0\setminus\{a\},F_i)+q(\{a\},F_i)-q(F_0,F_i)).$$

为了证明 $S\le \frac{1}{n}$, 我们只需证明局部不等式

$$q(F_0\setminus\{a\},F_i)+q(\{a\},F_i)-q(F_0,F_i))\le \frac{|F_i|}{n^2}.$$

将上式展开整理即

$$((e(F_0\setminus\{a\},F_i)-(|F_0|-1)e(\{a\},F_i))^2\le (|F_0|-1)|F_0||F_i|^2,$$

利用 $e(A,B)\le |A||B|$, 我们有 $|e(F_0\setminus\{a\},F_i)-(|F_0|-1)e(\{a\},F_i)|\le (|F_0|-1)|F_i|$, 即证得上式. 至此我们证完了 Regularity lemma. □

我们接下来介绍三个和 Regularity 相关的定理.

Theorem. (Graph Counting Theorem) 设图 H 的顶点为 $1,2,\cdots,k$. 图 G 的顶点集有子集 $V_1,V_2,\cdots,V_k$, 满足: 若 i,j 相连, 则 V_i,V_j 是 *ϵ-regular* 的. 独立随机选取 $v_i\in V_i$. 考虑事件 T: 若 i,j 相连, 则 v_i,v_j 相连. 则

$$|\mathbb{P}(T)-\prod_{i\sim j}d(V_i,V_j)|\le \epsilon e(H).$$

Proof. 我们对 $|H|$ 用归纳法. $|H|=0,1$ 时命题显然成立, 下设 $|H|\ge 2$. 由对称性, 不妨设 1,2 相连. 我们考虑事件 T': 若 i,j 相连, 且 $\{i,j\}\ne\{1,2\}$, 则 v_i,v_j 相连. 那么, 由归纳假设,

$$|\mathbb{P}(T')-\frac{\prod_{i\sim j}d(V_i,V_j)}{d(V_1,V_2)}|\le \epsilon(e(H)-1),$$

因此, 我们只需证明

$$|\mathbb{P}(T)-d(V_1,V_2)\mathbb{P}(T')|\le \epsilon.$$

由于 T,T' 的定义, 我们事实上只需证明固定 $v_3,\cdots,v_k$, v_1,v_2 随机选取时的上述不等式. 定义

$$A_1=\{v_1|v_1\in V_1,\text{对于一切}i>2,\text{若}i\text{与}1\text{相连},\text{则}v_i\text{与}v_1\text{相连}\},$$
$$A_2=\{v_2|v_2\in V_2,\text{对于一切}i>2,\text{若}i\text{与}2\text{相连},\text{则}v_i\text{与}v_2\text{相连}\}.$$

则固定 $v_3,\cdots,v_k$ 后, $p(T)$ 为从 V_1 中随机选一个元素, 它在 A_1 中, 从 V_2 中随机选一个元素, 它在 A_2 中, 且这两个元素相连的概率; $p(T')$ 为从 V_1 中随机选一个元素, 它在 A_1 中, 从 V_2 中随机选一个元素, 它在 A_2 中的概率. 即我们只需证明

$$|\frac{e(A_1,A_2)}{|V_1||V_2|}-d(V_1,V_2)\frac{|A_1||A_2|}{|V_1||V_2|}|\le \epsilon.$$

我们下面分情况讨论证明这个不等式. 若 $|A_1| \le \epsilon |V_1|$, 则

$$\frac{e(A_1, A_2)}{|V_1||V_2|} \le \frac{|A_1||A_2|}{|V_1||V_2|} \le \frac{|A_1|}{|V_1|} \le \epsilon,$$

且

$$d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|} \le \frac{|A_1|}{|V_1|} \le \epsilon,$$

因此命题得证; $|A_2| \le \epsilon |V_2|$ 时同理. 若 $|A_1| \ge \epsilon |V_1|, |A_2| \ge \epsilon |V_2|$, 由于 V_1, V_2 为 ϵ-*regular* 的, 我们有

$$|\frac{e(A_1, A_2)}{|V_1||V_2|} - d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|}| = |d(A_1, A_2) - d(V_1, V_2)\frac{|A_1||A_2|}{|V_1||V_2|}| \le \epsilon.$$

因此该不等式得证. □

下一个定理是 Graph Counting Theorem 和 Regularity lemma 的综合运用.

Theorem. (Graph Removing Theorem) 给定图 H 和 $\epsilon > 0$. 则存在实数 δ, 满足: 对于任意具有 n 个顶点的图 G, 若其含 H 的个数不超过 $\delta n^{v(H)}$, 则可以从 G 中去掉至多 ϵn^2 条边, 使得剩下的图不含 H.

Proof. 我们首先待定 ϵ_0, δ. 我们总是可以不妨设 $\delta n^{v(H)} > 1$. 因此, 我们取 δ 充分小, 此时 n 充分大, 由 Regularity lemma, 就可以不妨设 $V(G)$ 存在 ϵ_0-*regular* 的划分 $\{V_0, V_1, V_2, \cdots, V_k\}$. 我们现在去掉 G 中如下的边:

- 所有 V_0 中的点连出的边, 这样的边至多 $\epsilon_0 n^2$ 条.
- V_i 中内部点连的边, 这样的边至多 $k\frac{n^2}{k^2} < \epsilon_0 n^2$ 条.
- 不 ϵ_0-*regular* 的对 (V_i, V_j) 之间连的边, 这样的边至多 $\epsilon_0 k^2 \frac{n^2}{k^2} = \epsilon_0 n^2$ 条.
- $d(V_i, V_j) < \frac{\epsilon}{4}$ 的对 (V_i, V_j) 之间连的边, 这样的边至多 $\frac{\epsilon}{4} k^2 \frac{n^2}{k^2} = \frac{\epsilon}{4} n^2$ 条.

因此, 我们可以选取 ϵ_0 使得删去的边不超过 ϵn^2 条.

接下来我们证明剩下的图中没有 H. 若剩下的图中含 H, 不妨设 H 的顶点 v_i 属于集合 T_i. (注意, T_i 是 $V_1, V_2, \cdots, V_k$ 之一, 但 T_i 可能相同.) 此时, 若 i, j 连边, 则 T_i, T_j 之间连了边, 根据我们删去的边的条件知, T_i, T_j 是 ϵ_0-*regular* 的, 且 $d(T_i, T_j) \ge \frac{\epsilon}{4}$. 因此, 对集合 T_i 使用 Graph Counting Theorem, 知

$$\mathbb{P}(T) \ge \prod_{i \sim j} d(T_i, T_j) - \epsilon_0 e(H) \ge (\frac{\epsilon}{4})^{e(H)} - \epsilon_0 e(H).$$

取 ϵ_0 使 $(\frac{\epsilon}{4})^{e(H)} - \epsilon_0 e(H) = t > 0$. 注意到事件 T 发生就诱导一个图 H, 因此图中 H 的个数

$\#(H)$ 就满足

$$\#(H) = t\prod_{i=1}^{k}|T_i| \geq t(\frac{1-\epsilon}{k})^k n^t.$$

而 $k \leq M(\epsilon_0)$ 对某个函数 M, 故 H 的个数大于等于 $\delta(\epsilon, H)n^t$. 结论得证. □

Graph Removing Theorem 的证明是经典的运用 Regularity lemma 的手法: 我们首先取一个满足 Regularity lemma 的分划 $\{V_0, V_1, V_2, \cdots, V_k\}$, 再去掉一些性质比较差的边, 剩下的图就会具有非常良好的结构. 我们将在下一节中再次看到这一技术的运用.

最后一个定理是在图 G 中寻找特定子图 H 的相关结论.

Theorem. (Graph Embedding Theorem) 对于任意 $d > 0$ 和 Δ, 存在 $\epsilon = \epsilon(d, \Delta)$ 和 $c = c(d, \Delta)$, 使得下述命题成立:

给定正整数 h, r 和图 H, 顶点为 $v_1, v_2, \cdots, v_h$, 色数不超过 r, 最大度数小于 Δ. 那么, 若图 G 的顶点集 $V_1, V_2, \cdots, V_r$ 满足: 元素个数均大于 ch, V_i, V_j 两两之间均 ϵ-*regular* 且 $d(V_i, V_j) \geq d$, 则 G 含 H.

Proof. 证明: 我们先待定 ϵ 和 c. 假设 H 的一个 r 染色将 H 的顶点分成集合 $U_1, U_2, \cdots, U_r$. 我们现在来找一个 U_i 到 V_i 的嵌入. 假设 H 的顶点 v_i 将被嵌入集合 $V_{\sigma(i)}$. 我们接下来逐步选出 v_i.

对于 $0 \leq j < i \leq h$, 记 $b_{j,i}$ 为 $v_1, \cdots, v_j$ 中与 v_i 相连的点的个数. 假设我们已经选出了 G 中的点 $x_1, x_2, \cdots, x_j$ 作为 $v_1, v_2, \cdots, v_j$. 对于 $i > j$, 记 Y_i^j 为此时 x_i 的可能的选择, 即 Y_i^j 为集合 $V_{\sigma(i)}$ 中满足下述条件的点 x 所构成的集合: 对于任意 $m \leq j$, 若 v_i 与 v_m 相连, 则 x 与 x_m 相连. 我们对 j 归纳证明, 我们总能保证 $|Y_i^j| \geq (d-\epsilon)^{b_{j,i}}|V_{\sigma(i)}|$.

$j = 0$ 时命题显然. 设 $j-1$ 时已经证明问题, 考虑 j 的情形. 首先, 若 v_i 与 v_j 不相邻, 我们取 $Y_i^j = Y_i^{j-1}$ 即可. 由对称性, 不妨设和 v_j 相邻的顶点为 $v_k, \cdots, v_h$. 我们先指出下述显然的引理:

Lemma. 若图 G 中, 顶点集 A, B 是 ϵ-*regular* 的, 且 $d(A, B) = d$. 若 $B_1 \subseteq B$ 满足 $|B_1| \geq \epsilon|B|$, 那么 A 中至多有 $\epsilon|A|$ 个元素, 在 B_1 中的邻居数少于 $(d-\epsilon)|B_1|$.

考虑 A 中在 B_1 中的邻居数少于 $(d-\epsilon)|B_1|$ 的元素构成的子集即显然证得引理. 回到原题. 我们首先使 $(d-\epsilon)^{\Delta} > \epsilon(\Delta+1)$. 因此, 对任意正整数 $k \leq i \leq h$, 在引理中分别取 $A = V_{\sigma(j)}$, $B = V_{\sigma(i)}$, $B_1 = Y_i^{j-1}$, 即知 $Y_{\sigma(j)}$ 中至多只有 $\epsilon|Y_i^{j-1}|$ 个点, 在 Y_i^{j-1} 中与不超过 $(d-\epsilon)|Y_i^{j-1}|$ 个点相连. 那么, 去掉这些点后, Y_j^{j-1} 中至少还剩

$$|Y_j^{j-1}| - \delta\epsilon|Y_{\sigma(j)}| \geq (d-\epsilon)|Y_{\sigma(j)}| - \delta\epsilon|Y_{\sigma(j)}| \geq \epsilon|Y_{\sigma(j)}| = \epsilon ch$$

个点. 取 $c = \frac{1}{\epsilon}$, 则 Y_j^{j-1} 中至少还剩 h 个点, 必定能选出一个作为 x_j. 此时, 由 x_j 的定义, 对于任意正整数 $k \leq i \leq h$, 就有 $|Y_i^j| \geq (d-\epsilon)|Y_i^{j-1}| \geq (d-\epsilon)^{b_{j,i}}|V_{\sigma(i)}|$.

综上所述, 结论得证. □