Szemeredi Regularity Lemma

Vladimir

4 февраля 2014 г.

1. Work

Лемма Семереди о регулярности - очень примечательное утверждение в первую очередь своей общезначимостью. Любой граф достаточно большого размера можно разбить на "почти случайные" подграфы.

Этот набросок предлагает одно из доказательств этого утверждения.

2. Formulation

Лемма о регулярности

$\forall \varepsilon > 0, \forall t > 0 \, \exists T = T(\varepsilon, t), T \in N : \forall G = (V, E) : |V| > T, G$ имеет ε-регулярное разбиение $(V_0, V_1 ... V_k)$, причём $t \leq k \leq T$.

Замечание: в этом наброске будет использоваться слово "разбиение" как в случае разбиения вершин графа, и в этом случае разбиение покрывает все его вершины, так и в случае разбиения вершин подграфа (или, что то же самое, подмножества вершин графа), и в этом случае оно покрывает только все вершины этого подграфа и никак не связано с остальными вершинами.

Обозначения

Прежде, чем доказывать лемму, надо объяснить обозначения.

- $G = (V, E)$ - означает граф на V вершинах и E рёбрах.
- По умолчанию считается, что $A \subset V, B \subset V$
- Связность пары $e(A, B)$ обозначает число рёбер с началом в A и концом в B. A, B - непересекающиеся подмножества вершин графа.
- Плотность пары (A, B) - величина, считающаяся по формуле:

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

  Для полносвязных пар $d = 1$, а для любых других $d \leq 1$.

- ε-регулярная пара (A, B) - это такая пара непересекающихся подмножеств V, что для любых довольно больших $X \subset A, Y \subset B, |X| \geq \varepsilon|A|, |Y| \geq \varepsilon|B|$ верно:

$$|d(A, B) - d(X, Y)| \leq \varepsilon$$

Неформально говоря, это значит, что связность подмножеств пары примерно такая же, как и самих множеств.

- Разбиение $V = (V_0, V_1...V_k)$ на попарно непересекающиеся части, называется равномерным, если $|V_1| = ... = |V_k|$. При этом V_0 - это "особое" множество, которое как бы не участвует в разбиении, а собирает $|V_0|$ отдельных частей (по одной вершине) в кучу.

- Если разбиение P' получено из P путём дробления частей P, то оно называется "уточнением" P. (под уточнение также подходит процесс перекидывания вершин в "особое" множество)

- Равномерное разбиение называется регулярным (в смысле, описанном выше), если его особое множество не очень большое ($|V_0| \leq \varepsilon|V|$), а почти все части (кроме εk^2) попарно ε-регулярны.

- Индекс пары (A, B) - это величина, задающаяся формулой:

$$q(A, B) = \frac{(|A||B|)}{n^2} d^2(A, B)$$

- Индекс пары разбиений $(\mathcal{A}, \mathcal{B})$ ($\mathcal{A}$ - разбиение A, $\mathcal{B}$ - разбиение B, A и B не пересекаются) задаётся формулой:

$$q(\mathcal{A}, \mathcal{B}) = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} q(A', B')$$

- Индекс одного разбиения:

$$q(\mathcal{P}) = \sum_{A, B \in \mathcal{P}} q(A, B)$$

Здесь под A и B подразумеваются либо V_i, кроме V_0, либо элементы V_0, взятые поодиночке.

В этой сумме $C^2_{k+|V_0|}$ элементов вида $q(A, B)$.

$d(A, B)$ не превышает 1. Значит $q(\mathcal{P}) \leq \frac{1}{n^2} \sum_{A \in \mathcal{P}, B \in \mathcal{P}} |A||B|$. Очевидно, что максимум правой части достигается, когда каждая часть разбиения состоит просто из одной точки и количество членов в сумме максимально. Значит $q(\mathcal{P}) \leq \frac{C_n^2}{n^2} \leq \frac{1}{2}$.

Идея доказательства:

Для начала возьмём случайное равномерно разбиение графа G (вероятно, какие-то отдельные вершины придётся сбросить в особое множество, чтобы обеспечить равенство мощностей частей).
Возможно, что это разбиение уже оказалось ε-регулярным. Тогда цель достигнута.
Если же оно оказалось нерегулярным, то (как будет показано в следующей лемме), его можно уточнить (подразбить) таким образом, чтобы регулярность как минимум не ухудшилась, а взвешенное среднее квадрата плотности между парами частей разбиения (индекс) увеличилось как минимум на константу. Затем получившееся разбиение можно итеративно уточнять. Поскольку взвешенное среднее не может превышать 1, а на каждом шагу оно увеличивается на константу, то процесс завершится за конечное число шагов, после которого мы получим искомое ε-регулярное разбиение.

Лемма 1

Пусть $G = (V, E)$ - граф. P - разбиение V

- $\forall A, B \in \mathcal{P}(A \neq B) \forall \mathcal{A}, \mathcal{B}$ - разбиений A, B $q(\mathcal{A}, \mathcal{B}) \geq q(A, B)$
- Если пара $(A, B), (A, B \in \mathcal{P})$ не является ε-регулярной, то найдутся такие разбиения $(\mathcal{A} = (A_1, A_2), \mathcal{B} = (B_1, B_2))$ что $q(\mathcal{A}, \mathcal{B}) \geq q(A, B) + \frac{\varepsilon^4 |A||B|}{n^2}$
- Если P' - уточнение P, то $q(P') \geq q(P)$.

Доказательство

Сконструируем случайную величину Z:

Будем выбирать случайные вершины: $a \in A$ и $b \in B$ равномерно. Затем возьмём части разбиений $A' \in \mathcal{A}, B' \in \mathcal{B}$ так что $a \in A', b \in B'$. (Это эквивалентно тому, что будем случайным образом брать части разбиений из распределения с весами, равными мощностям этих частей)

Звучит сложно, но на самом деле всё просто: берём случайным образом вершину, и вместе с ней тащим всю ту часть, к которой она принадлежит (A'). *Потом ещё раз* (B').

$$Z = d(A', B')$$

Посчитаем свойства Z:

- $$M[Z] = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} \frac{|A'||B'|}{|A||B|} d(A', B') = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} \frac{|A'||B'|}{|A||B|} \cdot \frac{e(A', B,)}{|A'||B'|} = d(A, B)$$

$$M^2[Z] = d^2(A, B) = \frac{n^2}{|A||B|} q(A, B)$$

- $$M[Z^2] = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} \frac{|A'||B'|}{|A||B|} d^2(A', B') = \frac{n^2}{|A||B|} \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} \frac{|A'||B'|}{n^2} d^2(A', B') =$$
$$= \frac{n^2}{|A||B|} q(\mathcal{A}, \mathcal{B})$$
- $D[Z] = M[Z^2] - M^2[Z] = \frac{n^2}{|A||B|}(q(\mathcal{A}, \mathcal{B}) - q(A, B))$
- Из неравенства Иенсена следует:
$$M[Z^2] \geq M^2[Z]$$
Подставив посчитанные свойства, получим доказательство первого пункта леммы.
- Третье утверждение леммы следует из первого и определения индекса разбиения.
- Второе утверждение докажем следующим образом:

  Посмотрим на дисперсию: $D[Z] = \frac{n^2}{|A||B|}(q(\mathcal{A}, \mathcal{B}) - q(A, B))$

  Оценим её с другой стороны. Поскольку известно, что пара (A, B) не регулярна, то найдутся подмножества $A' \subset A$, $B' \subset B$, такие что $|A'| \geq \varepsilon |A|, |B'| \geq \varepsilon |B|$ и $|d(A', B') - d(A, B)| \geq \varepsilon$. Сделаем разбиения $\mathcal{A} = (A', A \backslash A'), \mathcal{B} = (B', B \backslash B')$

  Из этого следует "ручная оценка" на дисперсию:
$$D[Z] > \frac{|A'||B'|}{|A||B|} \varepsilon^2 \geq \varepsilon^4$$
  Сопоставляя две формулы для дисперсии, получим второе утверждение леммы.

Зачем это нужно? Затем, что можно взять случайное равномерное разбиение. Возможно, оно окажется нерегулярным, но применив лемму к каждой части разбиения, можно его улучшить.

Об этом говорит

Лемма 2

Пусть $\varepsilon \leq 1/4, \mathcal{P} = V_0, V_1 ... V_k$ - равномерное разбиение $V, |V_i| = c, |V_0| \leq \varepsilon n$. Если само разбиение нерегулярно, то можно применить к нему "уточнение" до размера l, $k \leq l, \leq k4^k, |V_0'| \leq |V_0| + \frac{n}{2^k}$, так, что $q(\mathcal{P}') \geq q(\mathcal{P}) + \frac{1}{2}\varepsilon^5$

Процедура "уточнения" следующая:

Для каждой части разбиения V_i сконструируем $k-2$ разбиений, путём применения Леммы 1 к каждой паре $(V_i, V_j),\ i \neq j$.

Затем построим диаграмму Венна из всех этих разбиений, и каждый сегмент диаграммы и будем считать куском нового, уточнённого разбиения.

Всего часть разобьётся на не более, чем 2^{k-1} частей.

Объединим их все в разбиение $\mathcal{Q}$. Для него $q(\mathcal{Q}) \geq q(\mathcal{P}) + \frac{\varepsilon^5}{2}$

Всё бы неплохо, но оно неравномерное.

Обозначим $b = \lfloor c/4^k \rfloor$ и разобьём $\mathcal{Q}$ на части такого размера произвольным образом, выкидывая лишние вершины в особое множество.

Получившееся разбиение назовём $\mathcal{P}'$, оно и будет нужным.

Собственно, эта процедура и задаёт механизм построения разбиения.

Возьмём граф, ε, t, и разобьём его произвольным образом на t равных частей. Если разбиение не получилось равномерным, применим лемму 2, при этом число частей увеличится не очень сильно, и особое множество также сильно не вырастет. Если снова не получилось регулярное разбиение, повторим операцию.

Заметим, что число "регулярных пар" на каждом шаге неубывает, а индекс увеличивается. Число шагов конечно, потому что индекс не может быть больше $1/2$.

Значит, на каком-то шаге мы получим регулярное разбиение.

Лемма Семереди доказана.