12 Lemat Szemerédiego

Zacznijmy od zdefiniowania *gęstości* i *ϵ-regularności* pary rozłącznych podzbiorów X i Y wierzchołków grafu G, oraz *podziału ϵ-regularnego.*

Gęstością pary (X, Y) w grafie G, gdzie $X, Y \subseteq VG)$ i $X \cap Y = \emptyset$, nazywamy liczbę $d_G(X,Y) = \frac{e_G(X,Y)}{|X||Y|}$. Dla ustalonego $\epsilon > 0$, parę (X, Y) nazywamy ϵ-regularną, gdy dla wszystkich par (X', Y'), gdzie $X' \subseteq X$, $Y' \subseteq Y$, $|X'| \geq \epsilon|X|$, $|Y'| \geq \epsilon|Y|$, zachodzi nierówność

$$|d_G(X', Y') - d_G(X, Y)| \leq \epsilon.$$

Para ϵ-regularna przypomina dwudzielny graf losowy, gdzie krawędzie pojawiają się niezależnie z prawdopodobieństwem d, $0 < d < 1$. W takim grafie, z prawdopodobieństwem bardzo bliskim 1, wszystkie dostatecznie duże pary bedą miały gęstość zbliżoną do d. Stąd częto pary ϵ-regularne nazywa się pseudolosowymi.

Podział $V(G) = V_0 \cup V_1 \cup \cdots \cup V_k$ nazywamy ϵ-regularnym, gdy $|V_0| \leq \epsilon|G|$, $|V_1| = \cdots = |V_k|$, i, co najważniejsze, liczba tych par spośród (V_i, V_j), $1 \leq i < j \leq k$, które nie są ϵ-regularne, wynosi nie więcej niż ϵk^2. Zbiór V_0 pełni w tej definicji rolę pomocniczą i jest czasem nazywany śmietnikiem.

Lemat 9 (o regularności grafów, Szemerédi, 1978, D 7.4.1) *Dla każdego $\epsilon > 0$ i naturalnego m, istnieje liczba M taka, że każdy graf o co najmniej m wierzchołkach posiada podział ϵ-regularny $(V_0, V_1, ..., V_k)$ dla pewnego $m \leq k \leq M$.*

Warunek $k \leq M$ gwarantuje, że zbiory V_i, $i = 1, \ldots, k$, są duże, bo $|V_i| \geq (1 - \epsilon)|G|/M$. Warunek $k \geq m$ pozwala oszacować z góry liczbę krawędzi wewnątrz zbiorów V_i, bo $\sum_{i=1}^{k} e(V_i) < k\frac{|G|^2}{2k^2} \leq \frac{|G|^2}{2m}$. Tak więc, lemat ten mówi, że większość krawędzi każdego dużego grafu można rozbić na niewielką liczbę pseudolosowych podgrafów dwudzielnych.

Idea dowodu: Wprowadza się indeks podziału $q(\mathcal{P})$, który jest ograniczony z góry przez $1/2$. Ponadto indeks podpodziału nie może być mniejszy od indeksu wyjściowego podziału. Główny krok polega na pokazaniu, że jeśli dany podział $(V_0, V_1, ..., V_k)$ nie jest ϵ-regularny, to można znaleźć jego podpodział na co najwyżej $k4^k$ części, którego indeks jest większy o co najmniej $\epsilon^5/2$. Powtarzając tę operację nie więcej niż $1/\epsilon^5$ razy musimy więc w końcu otrzymać podział ϵ-regularny (bo w przeciwnym razie indeks przekroczyłby $1/2$, co jest niemożliwe).

Dowód: Oznaczmy $n = |G|$. Dla pary rozłącznych podzbiorów $A, B \subseteq V(G)$ zdefiniujmy $q(A,B) = \frac{|A||B|}{n^2} d^2(A,B)$. Dalej, dla podziału $\mathcal{A}$ zbioru A i podziału $\mathcal{B}$ zbioru B, niech $q(\mathcal{A},\mathcal{B}) = \sum_{A' \in \mathcal{A}, B' \in \mathcal{B}} q(A',B')$. W końcu, dla podziału $\mathcal{P} = (C_1, \ldots, C_k)$ zbioru $V(G)$,

$$q(\mathcal{P}) = \sum_{1 \le i < j \le k} q(C_i, C_j).$$

Zauważmy, że jeśli $|C_i| \le n/k$, to $q(\mathcal{P}) \le \binom{k}{2} k^{-2} < 1/2$.

Lemat 10 (D 7.4.2) *(i)* $q(\mathcal{A},\mathcal{B}) \ge q(A,B)$
(ii) Jeśli $\mathcal{P}'$ jest podpodziałem $\mathcal{P}$, to $q(\mathcal{P}') \ge q(\mathcal{P})$.

Dowód: (i) Na podstawie nierówności Cauchy-Schwarza,

$$q(\mathcal{A},\mathcal{B}) = \frac{1}{n^2} \sum_{i,j} \frac{e^2(A_i,B_j)}{|A_i||B_j|} \ge \frac{1}{n^2} \frac{(\sum_{i,j} e(A_i,B_j))^2}{\sum_{i,j} |A_i||B_j|} = \frac{1}{n^2} \frac{e^2(A,B)}{|A||B|} = q(A,B).$$

(ii) Niech $\mathcal{P} = (C_1, \ldots, C_k)$ i $\mathcal{C}_i = \mathcal{P}'[C_i]$. Na podstawie (i),

$$q(\mathcal{P}) = \sum_{i,j} q(C_i,C_j) \le \sum_{i,j} q(\mathcal{C}_i,\mathcal{C}_j) \le \sum_{i,j} q(\mathcal{C}_i,\mathcal{C}_j) + \sum_i q(\mathcal{C}_i) = q(\mathcal{P}'). \blacksquare$$

Poprzedni lemat mówił, że drobiąc podział nie można zmniejszyć indeksu q. Następny lemat pokazuje, że dzieląc w odpowiedni sposób parę nieregularną wartość indeksu rośnie.

Lemat 11 (D 7.4.3) *Jeśli para (C,D) nie jest ϵ-regularna, to istnieją podpodziały $\mathcal{C} = (C_1, C_2)$ zbioru C i $\mathcal{D} = (D_1, D_2)$ zbioru D takie, że*

$$q(\mathcal{C},\mathcal{D}) \ge q(C,D) + \epsilon^4 \frac{|C||D|}{n^2}.$$

Dowód: Niech $C_1 \subseteq C$, $D_1 \subseteq D$, $|C_1| \ge \epsilon|C|$, $|D_1| \ge \epsilon|D|$, $\eta = d(C_1,D_1) - d(C,D)$ i $|\eta| > \epsilon$. To znaczy, para (C_1,D_1) świadczy o ϵ-nieregularności pary (C,D). Oznaczmy $c = |C|$, $d = |D|$, $e = e(C,D)$, $C_2 = C \setminus C_1$, $D_2 = D \setminus D_1$, $c_i = |C_i|$, $d_i = |D_i|$, $e_{ij} = e(C_i,D_j)$, $i,j = 1,2$. Wtedy, na podstawie nierówności C-S, tym razem dla sumy trzech składników,

$$n^2 q(\mathcal{C},\mathcal{D}) = \frac{e_{11}^2}{c_1 d_1} + \sum_{i+j>2} \frac{e_{ij}^2}{c_i d_j} \ge \frac{e_{11}^2}{c_1 d_1} + \frac{(e - e_{11})^2}{(cd - c_1 d_1)}.$$

Po podstawieniu $e_{11} = \frac{c_1 d_1 e}{cd} + \eta c_1 d_1$ i podniesieniu do kwadratu, prawa strona staje się sumą sześciu wyrazów, z których dwa się wzajemnie redukują, jeden (dodatni) odrzucamy, i otrzymujemy w końcu oszacowanie

$$n^2 q(\mathcal{C}, \mathcal{D}) \geq \frac{e^2}{cd} + \eta^2 c_1 d_1 > \frac{e^2}{cd} + \epsilon^4 cd = n^2 q(C, D) + \epsilon^4 |C||D|. \blacksquare$$

Ostatni z serii lematów pokazuje jak, w przypadku podziału nieregularnego, zsumować przyrosty indeksu, wynikające z Lematu 11, po wszystkich parach nieregularnych.

Uwaga. Przy obliczaniu indeksu podziału ze śmietnikiem C_0, z przyczyn technicznych każdy element śmietnika traktujemy jak osobny (jednoelementowy) zbiór podziału.

Lemat 12 (D 7.4.4) *Niech* $0 < \epsilon \leq 1/4$ *i niech* $\mathcal{P} = (C_0, \ldots, C_k)$ *będzie podziałem* $V(G)$*, który nie jest* ϵ*-regularny. Wtedy istnieje liczba naturalna* l*,* $k \leq l \leq k4^k$ *oraz podpodział* $\mathcal{P}' = (C_0', \ldots, C_l')$ *podziału* $\mathcal{P}$ *taki, że* $|C_0'| \leq |C_0| + n2^{-k}$*,* $|C_1'| = \cdots = |C_l'|$ *oraz* $q(\mathcal{P}') \geq g(\mathcal{P}) + \epsilon^5/2$*.*

Dowód: Dla przejrzystości rozumowania, przyjmijmy naiwnie, że $C_0 = \emptyset$, i że w trakcie dowodu wszystkie dzielenia liczb naturalnych wychodzą bez reszty. Oznaczmy $c = |C_i| = n/k$, $i = 1, \ldots, k$. Dla każdej pary (C_i, C_j), która nie jest ϵ-regularna, niech $\mathcal{C}_{ij}$ będzie podziałem C_i a $\mathcal{C}_{ji}$ – podziałem C_j, których istnienie gwarantuje Lemat 11. Wtedy $q(\mathcal{C}_{ij}, \mathcal{C}_{ji}) \geq q(C_i, C_j) + \epsilon^4 c^2/n^2$. Dla pozostałych par przyjmijmy $\mathcal{C}_{ij} = \{C_i\}$ i $\mathcal{C}_{ji} = \{C_j\}$. Dla ustalonego i, podziały $\mathcal{C}_{ij}$, $j \neq i$, „szatkują" zbiór C_i na co najwyżej 2^{k-1} rozłącznych części. Niech $\mathcal{C}_i$ będzie tym podziałem zbioru C_i, tzn. $\mathcal{C}_i$ jest minimalnym (w sensie liczby klas) podziałem zbioru C_i, który jest podpodziałem każdego podziału $\mathcal{C}_{ij}$, $j \neq i$. Niech $\mathcal{C}$ będzie sumą wszystkich podziałów $\mathcal{C}_i$, tj. $\mathcal{C} = \bigcup_{i=1}^k \mathcal{C}_i$. Zauważmy, że $k \leq |\mathcal{C}| \leq k2^{k-1} < k2^k$.

Pokażemy teraz, że $q(\mathcal{C}) \geq q(\mathcal{P}) + \epsilon^5/2$. (Podział $\mathcal{C}$ nie jest obiecanym podziałem, bo jego klasy nie są równe; w końcowej części dowodu osiągniemy równość klasową drobiąc go na małe kawałeczki, co na podstawie Lematu 10(ii) nie obniży wartości indeksu.) Na podstawie Lematów 10 i 11 mamy

$$q(\mathcal{C}) \geq \sum_{1 \leq i < j} q(\mathcal{C}_i, \mathcal{C}_j) \geq \sum_{1 \leq i < j} q(\mathcal{C}_{ij}, \mathcal{C}_{ji}) \geq \sum_{1 \leq i < j} q(C_i, C_j) + \epsilon k^2 \frac{\epsilon^4 c^2}{n^2} = q(\mathcal{P}) + \epsilon^5.$$

W rzeczywistości, uwzględniając śmietnik, otrzymujemy tylko $\epsilon^5/2$ (bo wtedy $c \geq (n - \epsilon n)/k \geq \frac{3}{4}(n/k)$).

Teraz rozdrobnimy podział $\mathcal{C}$, dzieląc każdy jego zbiór na podzbiory mocy $\lfloor c4^{-k} \rfloor$. Nowy podział $\mathcal{P}'$ ma $l \le k4^k$ podzbiorów (dlaczego?). Powracając do realiów, może się zdarzyć, że nie wszystkie klasy podziału $\mathcal{C}$ mają moc podzielną przez d. Wtedy reszty dorzucamy do śmietnika, który jednak nie wzrośnie o więcej niż $\lfloor c4^{-k} \rfloor |\mathcal{C}| \le c4^{-k}k2^k \le n2^{-k}$. ■

Końcówka dowodu: Niech dane będą $0 < \epsilon < 1/4$ i $m \ge 1$. Oznaczmy przez $s = \epsilon^{-5}$ maksymalną liczbę iteracji głównego kroku dowodu – znajdowania podpodziału aktualnego podziału. Niech k będzie najmniejszą liczbą naturalną taką, że $k \ge m$ oaz $2^{k-1} \ge s/\epsilon$ i $M = \max\{f^{(s)}(k), 2k/\epsilon\}$, gdzie $f(x) = x4^x$. Na przykład,

$$f^{(2)}(x) = x4^x4^{x4^x}, \qquad f^{(3)}(x) = x4^{x+x4^x+x4^{x+x4^x}}, \qquad \dots.$$

Udowodnimy lemat z tym właśnie monstrualnym M.

Niech G będzie dowolnym grafem o $n \ge m$ wierzchołkach. Dla $n \le M$ konkluzja jest trywialna. Załóżmy więc, że $n > M$ i podzielny dowolnie $V(G) = C_0 \cup C_1 \cup \cdots \cup C_k$ tak, by $|C_1| = \cdots = |C_k|$ a $|C_0| < k$. Przypuśćmy, że ten podział nie jest ϵ-regualrny. Ponieważ $k \le \epsilon M/2 < \epsilon n/2$, to możemy stosować Lemat 12 otrzymując nowy, drobniejszy podział o większym indeksie i śmietniku. Powtarzamy ten krok aż do skutku. Jednak nawet po s iteracjach, wielkość śmietnika nie przekroczy liczby $k + sn2^{-k} \le \epsilon n/2 + \epsilon n/2 = \epsilon n$. Ostateczny, ϵ-regularny podział będzie mieć co najwyżej $f^{(s)}(k) \le M$ części. ■

13 Zastosowania Lematu Szemerédiego

13.1 Twierdzenie Erdősa-Stone'a (Rozdziały 7.1 i 7.5 podręcznika)

13.2 Krawędziowe liczby Ramseya (Twierdzenie 9.2.2 w podręczniku)