1 Wykład czwarty — wstęp do ekspanderów

Definicja 1.1. Dla grafu $G = (V, E)$ macierzą sąsiedztwa tego grafu $M(G)$ nazwiemy macierz, gdzie wiersze i kolumny indeksujemy elementami V, i $a_{vw} = 1$ dla $vw \in E$, a 0 wpp.

Definicja 1.2. Multigraf, graf z pętlami.

1.1 Ekspansja kombinatoryczna

Definicje: multigraf, graf z pętlami, graf d–regularny.

Definicja 1.3. Współczynnik ekspansji wierzchołkowej grafu do a (oznaczany $h_a^V(G)$) to

$$\inf_{S \subset V(G): |S| \leqslant a|V(G)|} (|N\{S\}| - |S|)/|S|.$$

Współczynnik ekspansji wierzchołkowej $h^V(G)$ bez parametru to $h_{1/2}^V(G)$.

Przypomnieć definicję $N\{S\}$. Skoro graf jest d—regularny, to to coś jest nieujemne (bo zbiór S wypuszcza $d|S|$ krawędzi, które wchodzą do co najmniej $|S|$ wierzchołków). Czyli w szczególności ekspansja wierzchołkowa grafu niespójnego to zero, zaś ogólnie ekspansja wierzchołkowa do a grafu nie może przekroczyć $(1 - a)/a$ z dokładnością co do asymptotycznie nieistotnych problemów z podzielnością (dla $|S| = a|V(G)|$).

Definicja 1.4. Współczynnik ekspansji krawędziowej grafu (ozn. $h^E(G)$) to

$$\inf_{S \subset V(G): |S| \leqslant |V(G)|/2} |E(S, \bar{S})|/|S|.$$

To też można by definiować dla a, ale tu chyba nie ma sensu (być może powyżej też nie ma, zobaczymy). To dla grafu niespójnego znowu jest zerem, a ogólnie nie przekracza d.

Fakt 1.5. Jako przykład — oblicz ekspansję wierzchołkową i krawędziową powiedzmy kliki K_{2n} (wierzchołkowa to 1, krawędziowa to n).

Fakt 1.6. Jeśli graf ma ekspansję wierzchołkową h, to ma ekspansję wierzchołkową h (innymi słowy, $h^E \geqslant h^V$).

Dowód. Weźmy zbiór S, on ma $(1 + h)|S|$ sąsiadów, czyli przynajmniej $h|S|$ sąsiadów poza S, czyli wychodzi z niego przynajmniej $h|S|$ krawędzi. □

W drugą stronę nie ma implikacji:

Fakt 1.7. Niech $G = K_{2n,2n}$. Wtedy $h^V(G) = 0$, $h^E(G) = n/2$.

Dowód. Do pierwszego bierzemy za S jedną ze stron grafu. Do drugiego niech w S będzie k wierzchołków z jednej i l z drugiej, wtedy $|E(S, \bar{S})| = k(n-l)+l(n-k) = (k+l)n-2kl/(k+l)$. Lewy składnik to $n|S|$, prawy maksymalizuje się dla $k = l = n/2$, co daje $n|S|/2$, cbdo. □

1.2 Motywacje: redukcja losowości i błądzenia losowe

1.2.1 Redukcja losowości

Dla ustalenia uwagi, skupmy się na następującej definicji algorytmu randomizowanego:

Definicja 1.8. Algorytm $A(x, y)$ dla języka L nazwiemy randomizowanym z jednostronnym błędem, jeśli dla wejścia x i y będącego ciągiem $m(|x|)$ losowych bitów (z jednostajnym rozkładem), jeśli $x \notin L$ to odpowiada NIE z prawdopodobieństwem 1, zaś jeśli $x \in L$ to odpowiada TAK z prawdopodobieństwem co najmniej $1 - p_A$ (czyli myli się z prawdopodobieństwem co najwyżej p_A, i potrafi tylko powiedzieć czasem NIE jeśli $x \in L$). Załóżmy dodatkowo, że $p_A \leqslant 1/4$.

Przykładem takiego algorytmu jest test Millera–Rabina czy Solovaya–Strassena, czy liczba jest pierwsza. W przypadku takich algorytmów, by mieć właściwie pewność, czy $x \in L$, powtarzamy go niezależnie k razy. Szansa, że nasz algorym nas oszukał jest wtedy co najwyżej p_A^k — maleje wykładniczo. Lecz, jeśli w jednym przebiegu nasz algorytm potrzebował $m = m(|x|)$ bitów losowych, to by uruchomić go k razy, potrzebuje km bitów. Bity prawdziwie losowe są dość drogie — w praktyce często bierze się je z obserwowania środowiska: ruchów myszką, temperatury procesora itd — lecz te źródła mają dość małą przepustowość. Zredukowanie więc liczby potrzebnych bitów ma sens.

Spójrzmy wpierw na taki pomysł:

1. Weźmy G — graf o 2^m wierzchołkach (etykietowanych ciągami m–bitowymi), d–regularny, i o dobrej ekspansji K. Nie chcemy go konstruować w pamięci całego: chcemy tylko, będąc w wierzchołku v umieć szybko przeiterować $N(v)$.

2. Wylosujmy y_0 — wierzchołek G.

3. Weźmy stałą c i puśćmy $A(x, y)$ dla każdego y, takiego że $d_G(x, y) \leqslant c$.

Losowość ograniczyliśmy tutaj do m, ale kosztem czasu: uruchamiamy A około d^c razy. Zobaczmy, co nam to daje. Weźmy $x \in L$ i niech $Z(x) = \{y : A(x, y) = 0\}$, czyli zbiór tych wierzchołków G, gdzie algorytm A daje złą odpowiedź dla x. Dla $y \in V(G)$ definiujemy $B_c(y) = \{y' : d_G(y, y') \leqslant c\}$, czyli kula o promieniu c. Niech $Z = \{y : B_c(y) \subset Z(x)\}$, to są te y, które, jeśli wylosujemy w naszym ulepszonym algorytmie, to źle odpowiemy. Mamy więc $B_c(Z) \subset Z(x)$, czyli $|Z| \leqslant 2^m/4$. Wobec tego, na mocy definicji ekspansji $|B_i(Z)| \geqslant K^i|Z|$ dla $0 \leqslant i \leqslant c$, czyli

$$|Z| \leqslant \frac{|Z(x)|}{K^c} \leqslant \frac{2^m}{4K^c}.$$

Czyli szansa porażki wynosi co najwyżej $1/(4K^c)$ — nam maleje wykładniczo z wzrostem c.

W poprzednim przykładzie poświęciliśmy jednak dużo czasu. Spójrzmy na taki pomysł:

1. Weźmy G — graf o 2^m wierzchołkach (etykietowanych ciągami m–bitowymi), d–regularny, i o dobrej ekspansji.

2. Wylosujmy y_1 — wierzchołek G i puśćmy $A(x, y_1)$.

3. Następnie $k - 1$ razy wylosujmy y_{i+1} — sąsiada y_i — i puśćmy $A(x, y_{i+1})$.

Intuicyjnie, to powinno nieźle działać. Jeśli G ma dobrą ekspansję, będziemy w ten sposób symulować losowanie niezależne kolejnych y_i. Widać, że tu też zredukowaliśmy liczbę bitów losowych: potrzebujemy teraz $m + (k - 1) \log_2 d$. Do precyzyjnego dowodu, że to dobrze działa, potrzebujemy jednak paru narzędzi.

1.2.2 Błądzenie losowe

Widać, że w redukcji losowości pojawiło się następujące błądzenie losowe: zaczynamy z jakiegoś wierzchołka i następnie za każdym razem losujemy jednego z d sąsiadów i do niego idziemy. Spróbujmy to sformalizować. Niech $p = (p_i)_{i=1}^n$ będzie rozkładem prawdopodobieństwa na wierzchołkach grafu G, tj. $0 \leqslant p_i \leqslant 1$ i $\sum_{i=1}^n p_i = 1$ (z prawdopodobieństwem p_i jesteśmy w wierzchołku i). Zróbmy jeden krok błądzenia losowego. Zauważmy, że teraz rozkład prawdopodobieństwa wynosi

$$\frac{1}{d}M(G)p.$$

Po k krokach będzie to

$$\left(\frac{M(G)}{d}\right)^k p.$$

Tym, którzy bardziej uważali na zajęciach z algebry liniowej, metod numerycznych czy równań różniczkowych może się teraz przypomnieć, że o zachowaniu takiego wyrażenia można coś wywnioskować, robiąc rozkład Jordana macierzy $M(G)$ i coś tutaj mają do rzeczy wartości własne $M(G)$. Zajmiemy się więc teraz analizą spektralną $M(G)$.

1.3 Przerwa spektralna i teoria algebraiczna

Przypomnijmy, że dla macierzy M mówimy, że v jest wektorem własnym M z wartością własną λ, jeśli $Mv = \lambda v$. Wartości własne, przypomnijmy, są pierwiastkami wielomianu $\det(M - \lambda I)$, a zatem jest ich dokładnie n (z krotnościami).

Ale: macierz $M(G)$ jest symetryczna. Przytoczmy twierdzenie z algebry liniowej. Nie będziemy go tu dowodzić (choć nie jest to bardzo trudne).

Twierdzenie 1.9. Niech M będzie macierzą symetryczną nad $\mathbb{R}$. Wówczas rozkłada się ona nad $\mathbb{R}$ na UDU^{-1}, gdzie U jest ortonormalna ($U^{-1} = U^T$, rzędy U to wektory wzajemnie prostopadłe, o długości 1, czyli wyznaczające bazę ortonormalną), zaś D jest diagonalna.

W szczególności, z tego twierdzenia wynika, że:

1. Wszystkie wartości właśne $M(G)$ są rzeczywiste i jeśli wartość własna λ pojawia się i razy, to ma i wektorów własnych. Oznaczmy $\lambda_1 \geqslant \lambda_2 \geqslant \ldots \lambda_n$ wartości własne $M(G)$.

2. Mamy bazę ortonormalną wektorów własnych $M(G)$, bo macierz diagonalna tak ma, a U to tylko ortonormalna zamiana bazy. Oznaczmy przez v_i wektor własnym dla λ_i w tej bazie.

3. Pierwsza wartość własna to $\max x^T Ax/x^T x$, druga to to samo maksimum, ale po wektorach prostopadłych do v_1, etc.

4. Z kolei jeśli chcemy szukać wartości własnych maksymalnych co do modułu, to rozpatrujemy $\|Ax\|/\|x\|$, tu znowu możemy schodzić na przestrzenie prostopadłe, lub równoważnie $\max |x^T Ax|/x^T x$.

5. Jeśli $v = \sum_{i=1}^n a_i v_i$, to $v^T Av = \sum_{i=1}^n \lambda_i a_i^2$.

Poniżej zamieszczamy (dla chętnych) szkice dowodów części poniższych faktów, w tym twierdzenia 1.9.

Dowód. • λ jest wartością własną $M \iff$ istnieje wektor v, że $Mv = \lambda v \iff$ istnieje wektor v, że $(M - \lambda I)v = 0 \iff$ macierz $M - \lambda I$ nie jest pełnego rzędu $\iff \det(M - \lambda I) = 0$.

- Niech v, w będą wektorami własnymi odpowiadającymi wartościom własnym $\lambda \neq \mu$. Wtedy $\lambda \langle v, w\rangle = \langle Av, w\rangle = \left\langle v, A^T w\right\rangle = \langle v, Aw\rangle = \mu \langle v, w\rangle$, skąd (jeśli $\lambda \neq \mu$) dostajemy $\langle v, w\rangle = 0$.

- Wartości własne są rzeczywiste: jakaś wartość własna istnieje (bo zasadnicze twierdzenie algebry), jeśli jest zespolona, to $A(v + iw) = (a + ib)(v + iw)$, gdzie v i w są wektorami rzeczywistymi, i któryś jest niezerowy. Sprzęgam, $A(v - iw) = (a - ib)(v - iw)$. To teraz policzmy $(v - iw)^T A(v + iw) = (a + ib)(v - iW)^T(v + iw) = (a + ib)(|v|^2 + |w|^2)$, ale z drugiej strony $(v - iw)^T A(v + iw) = (A(v - iw))^T(v + iw) = (a - ib)(v - iw)(v + iw) = (a - ib)(|v|^2 + |w|^2)$. Wobec tego $b = 0$.

- Jedyna ciut trudniejsza część jest taka, że w wypadku symetrycznym k–krotnej wartości własnej odpowiada podprzestrzeń wektorów własnych wymiaru k. Ja to umiem zrobić dowodząc, że macierz symetryczna się diagonalizuje ortogonalnie. Dowód leci tak — weźmy dowolną wartość własną λ i jej wektor własny v_1. Uzupełnijmy jakkolwiek v_1 do bazy ortonormalnej (zakładam, że v_1 jest znormalizowany), ustawmy te wektory jako kolumny macierzy V. Wtedy $Ve_1 = v_1$. Wobec tego $V^T AVe_1 = \lambda V^T v_1 = \lambda e_1$, czyli e_1 jest wartością własną $V^T AV$, czyli pierwsza kolumna tej macierzy to $(\lambda, 0, 0, \ldots, 0)$. Co więcej, $(V^T AV)^T = V^T A^T V = V^T AV$, czyli to jest macierz symetryczna. Wobec tego pierwszy wiersz wygląda tak samo, a poza tym mamy podmacierz wymiaru $(n - 1) \times (n - 1)$. Ta podmacierz jest symetryczna, i diagonalizujemy ją indukcyjnie. I już. Wobec tego mamy diagonalizację ortonormalną. Jako, że zarówno wielomian charakterystyczny, jak i wymiar podprzestrzeni własnej to niezmienniki podobieństwa macierzy, to wystarczy zauważyć, że fakt, którego dowodzimy jest prawdziwy w sposób trywialny dla macierzy diagonalnej.

- Jak mamy bazę ortonormalną wektorów własnych, to wyrażenie $x^T Ax$ pisze się jako $\sum \lambda_i a_i^2$, gdzie $x = \sum a_i v_i$, z czego wynika charakteryzacja kolejnych wartości własnych. Druga wynika z tego, że $\|Ax\| = \sqrt{\langle Ax, Ax\rangle} = \sqrt{\sum \lambda_i^2 a_i^2}$ dla wektora normalnego x.

□

Dla multigrafów nieskierowanych $M(G)$ jest symetryczna. Jeśli G jest d–regularny, to suma elementów w każdym wierszu i w każdej kolumnie $M(G)$ jest równa d.

Fakt 1.10. d jest wartością własną $M(G)$ odpowiadającą wektorowi $(1, 1, \ldots, 1)$. To jest największa co do modułu wartość własna $M(G)$ (to dałem na ćwiczenia).

Dowód. Że jest wartością własną, to wystarczy przemnożyć, wynika z regularności. Że jest największa co do modułu to łatwo sprawdzić, że pierwsza norma wektora po pomnożeniu przez A rośnie co najwyżej d razy. □

Definicja 1.11. Względną przerwą spektralną $\Delta(G)$ grafu d–regularnego G nazwiemy liczbę $d - \lambda_2$, gdzie λ_2 jest drugą wartością własną macierzy $M(G)$. Bezwzględną przerwą spektralną $\Delta'(G)$ nazwiemy $d - \max(|\lambda_2|, |\lambda_n|)$.

Uwaga — (zapewne za dwa wykłady) udowodnimy, że to faktycznie ma związek z ekspansją. Przykład:

Fakt 1.12. Graf niespójny ma względną i bezwzględną przerwę spektralną 0 (a także, oczywiście, ekspansję wierzchołkową i krawędziową zero).

Dowód. Niech $V_1 \subset V$ będzie zbiorem wierzchołków pewnej spójnej składowej G. Wtedy $M(G)\mathbf{1}_{V_1} = d\mathbf{1}_{V_1}$, a zatem 1 jest podwójną wartością własną M. □

Fakt 1.13. Klika ma względną i bezwzględną przerwę spektralną $d = N - 1$.

Dowód. Pierwszą wartością własną jest oczywiście $N - 1$. Jeśli wezmę dowolny wektor prostopadły do $(1, 1, \ldots, 1)$, to $Av = 0$, skąd v jest wektorem własnym z wartością własną zero — czyli przestrzeń tych wektorów ma wymiar $n - 1$, a zatem wszystkie pozostałe wartości własne są zerami. □

1.4 Połączenie teorii — błądzenia losowe

Mając w ręku narzędzia spektralne, możemy zanalizować błądzenie losowe. Jak już zauważyliśmy wczesniej, zaczynając od rozkładu p, po k krokach będziemy mieli rozkład prawdopodobieństwa

$$p^{(k)} = \left(\frac{M(G)}{d}\right)^k p.$$

Skorzystajmy z rozkładu $M(G) = UDU^{-1}$ i niech $p = \sum_{i=1}^{n} a_i v_i$. Wówczas

$$p^{(k)} = \sum_{i=1}^{n} v_i \left(\frac{\lambda_i}{d}\right)^k a_i.$$

Co oznacza, że współczynniki $p^{(k)}$ przy v_i dla $i \neq 1$ wygasają wykładniczo (i tym szybciej, im większa jest bezwzględna przerwa spektralna), a ten przy v_1 zostaje. Czyli $p^{(k)}$ zbiega wykładniczo do v_1, czyli rozkładu jednostajnego. Bezwzględna przerwa spektralna rządzi prędkością zbiegania.

1.5 Połączenie teorii — dowody zależności ekspansji kombinatorycznych i spektralnych

Teraz zbadamy zależności pomiędzy czterema zdefiniowanymi pojęciami ekspansji. Generalnie względna przerwa spektralna odpowiada ekspansji krawędziowej, a bezwzględna — wierzchołkowej. W szczególności $h^V(G) \leqslant h^E(G)$ oraz $\Delta'(G) \leqslant \Delta(G)$.

(większość z tego, co poniżej, przeleci na szósty wykład)

Teraz udowodnimy dwa twierdzenia, pierwsze wiążące $\Delta(G)$ z $h^E(G)$, a drugie — $\Delta'(G)$ z $h^V(G)$.

Twierdzenie 1.14. Niech G będzie grafem d–regularnym. Wtedy

$$\Delta(G)/2 \leqslant h^E(G) \leqslant \sqrt{2d\Delta(G)}.$$

Wpierw może komentarz — kluczowe w tych stwierdzeniach jest to, że nie ma zależności od N — czyli jeśli mamy jakąś ekspansję krawędziową, to dostaniemy jakąś względną przerwę spektralną, i na odwrót.

Dowód. Wpierw prosty kierunek — że względna przerwa spektralna daje ekspansję krawędziową. Załóżmy, że mamy względną przerwę spektralną Δ. Weźmy jakiś zbiór S. Wpierw zauważmy, że jeśli weźmiemy wektor $\mathbf{1}_S$, to $\mathbf{1}_S^T A \mathbf{1}_{S'}$ to liczba krawędzi łączących S i S'. A tymczasem pierwsza wartość własna to supremum napisów $x^T Ax/x^T x$, zaś druga to supremum tych samych napisów po x prostopadłych do $(1, 1, \ldots, 1)$, czyli sumujących się do zera. Zróbmy zatem wektor sumujący się do zera: kładziemy $x = \mathbf{1}_S/|S| - \mathbf{1}_{\bar{S}}/|\bar{S}|$. Wtedy wiemy, że $(d - \Delta) \geqslant (x^T Ax)/(x^T x) = \left(E(S,S)/|S|^2 - 2E(S,\bar{S})/|S||\bar{S}| + E(\bar{S},\bar{S})/|\bar{S}|^2\right)/(\frac{1}{|S|} + \frac{1}{|\bar{S}|}$.

Wyliczamy $E(S,S) = (d|S| - E(S,\bar{S}))$ i $E(\bar{S},\bar{S}) = (d|\bar{S}| - E(S,\bar{S}))$, wstawiamy, dostajemy $(d-\Delta) \geqslant \Big(d/|S| + d/|\bar{S}| - E(S,\bar{S})(1/|S|^2 - 2/|S||\bar{S}| + 1/|\bar{S}|^2)\Big)/(1/|S| + 1/|\bar{S}|)$. Niech $c = 1/|S| + 1/|\bar{S}|$, to dostajemy $(d-\Delta) \geqslant (cd - E(S,\bar{S})c^2)/c = d - cE(S,\bar{S})$. Teraz zauważmy, że $|S|c = 1 + \frac{|S|}{|\bar{S}|} \leqslant 2$ dla $|S| \leqslant |V(G)|/2$, skąd mamy $E(S,\bar{S})/|S| \geqslant \Delta/2$. Biorąc po lewej infimum, dostaję $h^E(G) \geqslant \Delta/2$. To jest HLW 36.

Teraz w drugą stronę, to jest nieco trudniejszy kierunek. Weźmy macierz A, niech x będzie wektorem własnym odpowiadającym wartości własnej λ_2. Niech $\bar{x}$ będzie równe x tam, gdzie współrzędne x są dodatnie, i zero wszędzie indziej. Zakładamy, że $\bar{x}$ ma co najwyżej $|V(G)|/2$ niezerowych współrzędnych, w przeciwnym razie możemy spojrzeć na $-x$. Spójrzmy na $\sum_{vw\in E} |\bar{x}_v^2 - \bar{x}_w^2|$ (ta wielkość pozwala powiązać kombinatoryczne własności ekspansji z własnościami spektralnymi, warto ją zapamiętać).

Z jednej strony możemy to powiązać z ekspansją krawędziową następująco: uporządkujmy wierzchołki $v_1, v_2, \ldots, v_n$ tak, by $x_{v_1} \geqslant x_{v_2} \geqslant \ldots \geqslant x_{v_n}$. Niech $i \leqslant j$, przedstawmy $|\bar{x}_{v_i}^2 - \bar{x}_{v_j}^2| = \sum_{k=i}^{j-1} \bar{x}_{v_k}^2 - \bar{x}_{v_{k+1}}^2$. Teraz różnica $\bar{x}_{v_k}^2 - \bar{x}_{v_{k+1}}^2$ dla konkretnego k pojawia się w tej sumie raz dla każdej krawędzi łączącej $\{v_1, v_2, \ldots, v_k\}$ z $\{v_{k+1}, \ldots, v_n\}$. Ta różnica jest zerowa dla $k \geqslant n/2$, a dla mniejszych k możemy skorzystać z ekspansji krawędziowej, i powiedzieć, że tych krawędzi jest minimum $kh^E(G)$. Stąd $\sum_{vw\in E} |\bar{x}_v^2 - \bar{x}_w^2| \geqslant \sum_k kh^E(G)(\bar{x}_k^2 - \bar{x}_{k+1}^2) = h^E(G)\sum_k \bar{x}_k^2 = h^E(G)\|\bar{x}\|^2$.

Z drugiej strony możemy też tę wartość powiązać z przerwą spektralną. Otóż $\sum_{vw\in E} |\bar{x}_v^2 - \bar{x}_w^2| = \sum_{vw\in E} |\bar{x}_v - \bar{x}_w||\bar{x}_v + \bar{x}_w| \leqslant \sqrt{\sum |\bar{x}_v - \bar{x}_w|^2}\sqrt{\sum |\bar{x}_v + \bar{x}_w|^2}$ z nierówności Cauchy–Buniakowskiego–Schwarza. Każdy z fragmentów szacujemy oddzielnie.

Drugi jest prostszy: $\sqrt{\sum |\bar{x}_v + \bar{x}_w|^2} \leqslant \sqrt{\sum 2(\bar{x}_v^2 + \bar{x}_w^2)} = \sqrt{2d}\|\bar{x}\|$, bo skoro sumujemy po krawędziach, to każde $\bar{x}_v^2$ pojawia się dokładnie d razy w sumie.

Pierwszy jest też prosty: rozważmy $\sum |\bar{x}_v - \bar{x}_w|^2$, i otwórzmy nawiasy. To pojawi się d razy każdy $\bar{x}_v^2$, to już ćwiczyliśmy, oraz $-2\sum_{vw\in E} \bar{x}_v\bar{x}_w = -\sum_v \bar{x}_v \sum_w a_{vw}\bar{x}_w$ (dwójka znika, bo każdy iloczyn pojawia się raz przy sumowaniu po v, a raz po w). Czyli $\sum |\bar{x}_v - \bar{x}_w|^2 \leqslant d\|\bar{x}\|^2 - \sum_v \bar{x}_v \sum_w a_{vw}\bar{x}_w \leqslant d\|\bar{x}\|^2 - \sum_v \bar{x}_v \sum_w a_{vw}x_w$ — liczby x_w tam, gdzie różnią się od $\bar{x}_w$, to są ujemne, mnożymy przez coś dodatniego i odejmujemy, czyli całość wzrasta po zamianie. Stąd $= d\|\bar{x}\|^2 - \sum_v \bar{x}_v \sum_w a_{vw}x_w = d\|\bar{x}\|^2 - \sum_v \bar{x}_v\lambda_2 x_v = d\|\bar{x}\|^2 - \lambda\|\bar{x}\|^2$ (bo tam, gdzie mamy ujemne rzeczy w x_v, to mnożymy przez zera w $\bar{x}_v$). Stąd $\sqrt{\sum \|\bar{x}_v - \bar{x}_w\|^2} \leqslant \|\bar{x}\|\sqrt{\Delta}$.

Łączymy otrzymane nierówności w jedną, i dostajemy $h(G)\|\bar{x}\|^2 \leqslant \|\bar{x}\|^2\sqrt{2d}\sqrt{\Delta(G)}$, dzielimy stronami przez $\|\bar{x}\|^2$ i dostajemy tezę. □

W HLW 36–37 są przykłady, że to jest ciasne. Oni twierdzą, że jedno ciśnienie daje kostka, a drugie ciśnienie daje cykl. Zrobimy na ćwiczeniach.

Definicja 1.15. Kwadratem grafu G nazwiemy graf, którego macierzą sąsiedztwa jest $M^2(G)$. Krawędzie w G^2 odpowiadają ścieżkom długości dwa w grafie G.

Fakt 1.16. $h^V(G) \leqslant h^V(G^2)$.

Dowód. Weźmy dowolny zbiór $S \subset V(G^2) = V(G)$. Wtedy w G ma on przynajmniej $(1 + h^V(G))|S|$ sąsiadów. Weźmy dowolny $T \subset N_G\{S\}$, gdzie $|T| = |S|$ (przypomnijmy, ekspansja wierzchołkowa jest zawsze nieujemna, więc się da). I teraz z ekspansji dla G mamy $|N_G\{T\}| \geqslant (1 + h^V(G))|T| = (1 + h^V(G))|S|$, a $N_G\{T\} \subset N_G\{N_G\{S\}\} = N_{G^2}\{S\}$. □

Twierdzenie 1.17. Niech G będzie grafem d–regularnym. Wtedy

$$\frac{d^2 - (d - \Delta'(G))^2}{d^2 + (d - \Delta'(G))^2} \leqslant h^V(G) \leqslant d\sqrt{2}\sqrt{d^2 - (d - \Delta'(G))^2}.$$

Dowód. Będziemy rozważać związki między $\|v\|$ a ekspansją. Tu nie trzeba robić tej sztuczki z odejmowaniem, bo interesuje nas druga co do modułu wartość własna, czyli po prostu ekstremalizujemy $\|Av\|/\|v\|$ po $\sum v_i = 0$. Zauważmy, że $\|v\|^2 \geqslant 1/|\mathrm{supp}v|$ dla $\|v\|_1 = 1$ (to jest nierówność CBS: $1 = \|v\|_1 \leqslant \sqrt{|\mathrm{supp}v|}\|v\|$). No to weźmy dowolny wektor $v = \mathbf{1}_S/|S|$. Przepuszczamy go przez macierz A, dostajemy Av, i $\|Av\|^2 \geqslant d^2/|\mathrm{supp}Av| = d^2/|\Gamma(S)|$, zaś $\|v\|^2 = 1/|S|$. Teraz każdy z tych wektorów po odjęciu n będzie prostopadły do n, czyli $1/|S| = \|v\|^2 = \|n\|^2 + \|v'\|^2 = 1/N + \|v'\|$, oraz $d^2/|\Gamma(S)| \leqslant \|Av\|^2 = \|An + Av'\|^2 = \|dn + Av'\|^2 = d^2\|n\|^2 + \|Av'\|^2 \leqslant d^2/N + \lambda^2\|v'\|^2$. Dzielimy drugą równość przez d^2 i wstawiamy $\|v'\|$ z pierwszej, dostajemy $1/|\Gamma(S)| \leqslant 1/N + (\lambda/d)^2(1/|S| - 1/N)$. Teraz $N \geqslant 2|S|$, i $1/N$ jest po prawej z dodatnim współczynnikiem, czyli $1/|\Gamma(S)| \leqslant (\lambda/d)^2/|S| + (1 - (\lambda/d)^2)/2|S|$, co daje $|\Gamma(S)| \geqslant |S|\frac{2}{(\lambda/d)^2+1}$, czyli $h^V(G) \geqslant \frac{d^2-(d-\Delta')^2}{d^2+(d-\Delta')^2}$.

To teraz jeśli mamy ekspansję wierzchołkową h dla G, to mamy wierzchołkową h dla G^2 z faktu powyżej, to mamy krawędziową h dla kwadratu G^2 (bo wierzchołkowa jest nie większa niż krawędziowa) to mamy względną przerwę dla kwadratu (z twierdzenia poprzedniego), to mamy bezwzględną przerwę dla grafu pierwotnego (bo wartości własne kwadratu macierzy to kwadraty wartości własnych macierzy). I już, dostajemy, mam wrażenie, że $\Delta'_G \geqslant d - \sqrt{d^2 - \frac{(h^V)^2}{2d^2}}$. To nie jest najlepszy możliwy wynik, bo jak mamy ekspansję wierzchołkową, to da się wydusić lepszą ekspansję spektralną niż tylko przechodząc przez krawędzie, ale jest akceptowalny, bo dowodzi, że jak jest ekspansja wierzchołkowa niezależna od N, to jest i spektralna niezależna od N. □

To powyżej z góry odpowiada szacowaniu $\Delta'(G) \geqslant d - \sqrt{d^2 - \frac{h^V(G)^2}{2d^2}}$. Da się wyciągnąć lepiej:

Twierdzenie 1.18. $\Delta'(G) \geqslant d - \sqrt{d^2 - \frac{h^V(G)^2}{8+4h^V(G)^2}}$.

(ten dowód jest w DH).

2 Wykład piąty — zygzak

Przypomnijmy sobie pokrótce poprzedni wykład. Wprowadziliśmy pojęcia ekspansji wierzchołkowej, krawędziowej, oraz dwóch przerw spektralnych, pokazaliśmy, że odpowiednia przerwa spektralna wiąże się z odpowiednią ekspansją. Tu może podam jeszcze jeden wynik, który nie zmieścił się na poprzednim wykładzie — Expander Mixing Lemma:

Twierdzenie 2.1 (Expander Mixing Lemma). Niech d będzie grafem d–regularnym o n wierzchołkach, i $\lambda = d - \Delta'(G)$. Wtedy dla każdych $S, T \subset V$, zachodzi

$$\left||E(S,T)| - \frac{d|S||T|}{n}\right| \leqslant \lambda\sqrt{|S||T|}.$$

To mówi, że im większa przerwa spektralna (bezwzględna), tym bardziej graf przypomina graf losowy — po lewej stronie mamy różnicę między faktyczną liczbą krawędzie między S i T, a tym, czego oczekujemy po grafie losowym. Dowód pewnie będzie na ćwiczeniach, ale spisuję go dla porządku.

Przypomnijmy, że motywacją tego sportu było zastosowanie w algorytmach losowych. Idea — choćby ta najprostsza — była taka, że biorąc d^c wierzchołków otrzymujemy — bez dodatkowego marnowania losowych bitów — algorytm, który myli się z prawdopodobieństwem najwyżej $\frac{1}{4(1+h_V)^c}$, a działa w czasie d^c. Stosowaliśmy do tego ekspander na 2^N wierzchołkach, o stopniu d i ekspansji h_V.

Żeby ta idea miała sens, to musimy być w stanie wykonstruować ekspander, który ma dowolnie wiele wierzchołków, a stały stopień i oddzieloną od zera ekspansję. Co więcej (choć tym będziemy się zajmować dziś ciut mniej) potrzebujemy, by był on lokalnie konstruowalny — tj. by w rozsądnym czasie (np. $Poly(d)Polylog(N)$) dało się określić sąsiedztwo wierzchołka, bez konstruowania całego grafu (bo graf na 2^N wierzchołkach, należy założyć, nie mieści nam się w pamięci, a nawet gdyby, to jeśli stać nas czasowo na jego konstrukcję, to stać nas na algorytm, który w ogóle nie błądzi).

Cel — rodzina grafów o rosnącej mocy, stałym stopniu i stałej niezerowej ekspansji. Generalnie łatwo jest poprawiać ekspansję eksplodując stopień (najprościej poprzez kwadratowanie). Chodzi jednak o to, żeby utrzymać stały stopień, stałą ekspansję, a iść z rozmarem do nieskończoności.

Narzędzie — produkt grafów, dużego i małego, którego rozmiar to produkt rozmiarów, stopień idzie od małego, a ekspansja jest dobra jeśli obydwie tamte były dobre.

Formalnie własności:

Twierdzenie 2.2. Niech G i H będą grafami, przy czym $\deg G = |H|$. Wtedy ich produkt zygzakowy (zygzak) $G \circledcirc H$ ma następujące właściwości:

- $|G \circledcirc H| = |G||H|$;
- $\deg G \circledcirc H = (\deg H)^2$;
- Jeśli druga wartość własna co do modułu G to $\alpha \deg G$, zaś druga wartość własna co do modułu H to $\beta \deg H$, to druga wartość własna $G \circledcirc H$ jest nie większa co do modułu niż $(\alpha + \beta + \beta^2) \deg^2 H$.

Znajomość tych własności (oraz wprowadzonych poprzednio konceptów tensorowania i kwadratu) wystarcza do pokazania, jak zaczynając od ustalonego grafu zrobić dobry ekspander. Zaczynamy od ustalonego grafu H o d^8 wierzchołkach, stopniu d oraz bezwzględnej przerwie spektralnej Δ'. Konstruujemy ciąg grafów G_t, gdzie G_t ma d^{8t} wierzchołków, stopień d^2 oraz bezwzględną przerwę Δ'_t. $G_1 = H^2$, $G_2 = H \times H$, zaś

$$G_t = \left(G_{\lfloor \frac{t-1}{2} \rfloor} \times G_{\lceil \frac{t-1}{2} \rceil}\right)^2 \circledcirc H.$$

Twierdzenie 2.3. Graf G_t ma rozmiar d^{8n} i stopień d^2. Jeśli druga wartość własna co do modułu H to nie więcej niż $d/5$, to druga wartość własna co do modułu G_n to nie więcej niż $2d^2/5$.

Mamy zatem, zwróćmy uwagę, ciąg grafów o dowolnie dużej mocy, stałym stopniu oraz stałej (i dodatniej) przerwie spektralnej, to oczywiście przekłada się na stałą i dodatnią (bo d^2 jest stałe) ekspansję.

Dowód. Wpierw stopień. $G_{t-1/2}$ ma z założenia stopień d^2, iloczyn tensorowy ma stopień d^4, kwadrat — stopień d^8, to oznacza, że mogę zygzakować, dostaję stopień zygzaku d^2.

Moc — iloczyn tensorowy ma moc $d^{8\lceil t-1/2 \rceil} d^{8\lfloor t-1/2 \rfloor} = d^{8(t-1)}$, kwadrat nie zmienia mocy, zygzak domnaża brakujące d^8.

I teraz bezwzględna przerwa spektralna. Druga co do modułu wartość własna $G_{t-1/2}$ to co najwyżej $2d^2/5$, tensora — $2d^4/5$, kwadratu — $4d^4/25$, a zatem zygzaka, z własności, co najwyżej $2d^2/5$, czyli sukces. □

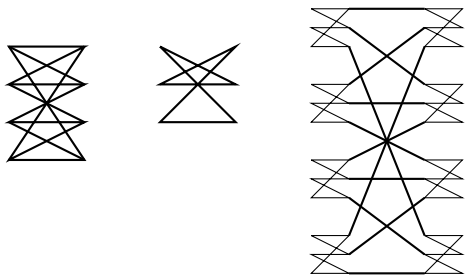
W tym dowodzie warto zwrócić uwagę na trzy rzeczy. Po pierwsze — zobaczmy, za co odpowiadają poszczególne operacje w konstrukcji indukcyjnej. Kwadrat odpowiada za poprawienie

ekspansji, niestety robi to kosztem powiększenia stopnia. Zygzak, kosztem kontrolowanego pogorszenia ekspansji, poprawia stopień z powrotem, te dwie operacje bilansują się wzajemnie. Można by ograniczyć się do tych dwóch operacji, ale wtedy przyrost rozmiaru grafu byłby dość powolny — tensorowanie odpowiada za istotne przyspieszenie go.

Tu uwagi algorytmiczne — jeśli dla grafów G i H pytanie o sąsiadów danego wierzchołka v kosztuje nas czas C, to dla G^2 by dostać sąsiedztwo wykonujemy $d+1$ zapytań o sąsiedztwo, dla $G \times H$ wystarczy jedno pytanie o sąsiedztwo w G i jedno w H, do zygzaku — jak zobaczymy — potrzeba $d+1$ pytań o sąsiedztwo w G i d_H^2 pytań o sąsiedztwo w H. W naszym wypadku wszystkie stopnie są ograniczone przez stałą, czyli należy myśleć, że na każdym poziomie zadajemy D pytań o sąsiedztwa. Aby odpowiedzieć na każde z nich, znowu musimy zadać D pytań o sąsiedztwa, i tak wgłąb — konkretniej, aby odpowiedzieć na jedno pytanie na poziomie t potrzebujemy D odpowiedzi na poziomie $t/2$. Łącznie zatem zadamy $D^{\log t} = t^{\log_D}$ pytań na poziomie pierwszym (na które odpowiadamy w czasie stałym), a t, przypomnijmy, jest logarytmiczne w rozmiarze naszego pierwotnego grafu. Czyli faktycznie na pytanie o sąsiedztwo odpowiadamy w czasie polylog, czyli akceptowalnym. Przykładowo w naszym zastosowaniu graf miał 2^N wierzchołków, gdzie N to liczba bitów losowych potrzebnych pierwotnemu algorytmowi.

Trzecia rzecz — może nas zaniepokoić nieco, skąd tak właściwie mamy wziąć ten graf startowy. Otóż jest kilka technologii do dyspozycji, ja zaproponuję dwie. Po pierwsze, na ćwiczeniach pokażemy, jak robić stosunkowo przyzwoite ekspandery — co prawda, żeby poprawiać ekspansję, będziemy musieli w nich zwiększać stopień, ale w sposób raczej umiarkowany, i uda nam się dostać ekspandery bazowe spełniające to, co trzeba. Druga technologia jest taka, że być może udowodnimy na ćwiczeniach, że odpowiednio dobrany dostatecznie duży losowy graf ma stopień 3, dowolnie duży rozmiar i ograniczoną niezależnie od rozmiaru stałą ekspansji (przez, powiedzmy, 10^{-3}) z dodatnim prawdopodobieństwem. Wobec tego aby dostać ekspander bazowy taki losowy graf podniesiemy do potęgi $-1/\log_5(999/1000)$, to dostaniemy ekspander o dobrej stałej ekspansji i ograniczonym stopniu. A taki losowy graf możemy znaleźć choćby bezpośrednim, siłowym przeszukaniem w stałym czasie.

Definicja zygzaka — idea oraz definicja prawie poprawna. Chcemy korzystać z tego, że moc H to stopień G — czyli wierzchołki H możemy traktować jako etykiety krawędzi w G. Robimy tak, że dla każdego wierzchołka (zarówno w G jak i w H) etykietujemy krawędzie wychodzące z danego wierzchołka liczbami od jedności do stopnia danego grafu. Wobec tego chcemy zdefiniować krawędź (i, j) (gdzie i, j to etykiety krawędzi w H) z wierzchołka (s, t), gdzie $s \in V(G)$, $t \in V(H)$. To wpierw wykonujemy krok w H wedle etykiety i — dostajemy $(s, t[i])$. Teraz wykonujemy krok w G używając aktualnego wierzchołka w H jako etykiety — dostajemy $s[t[i]], t[i]$. I teraz kroczymy dalej, wykonujemy krok znowu w grafie H, wedle etykiety j, dostajemy $(s[t[i]], t[i][j])$.



Rysunek 1: Zygzak grafu o czterech i grafu o trzech wierzchołkach (dla wygody rysunku każdy graf jest rozrzucony na wierzchołki wejściowe i wyjściowe)

Definicja powyżej jest prawie dobra, ma tylko jeden problem — nie daje grafu nieskierowa-

nego. Problem jest taki, że niestety dana krawędź może być w dwie różne strony etykietowana dwoma różnymi liczbami (i może nie dać się z tym nic zrobić, np. dla trójkąta, więcej w tym temacie będzie na wykładach o kolorowaniach). A na chwilę obecną niestety trzeba zrobić tak, że idąc po grafie G jednocześnie podmieniamy etykietę (czyli wierzchołek z H) na "etykietę odwrotną" — czyli tę etykietę, którą trzeba pójść, żeby wrócić się po naszej krawędzi.

Wpierw zróbmy przykład, jak to ma działać: rysujemy graf $H = 1234, \{12, 13, 34, 24\}$ i $G = 123456$, w którym niepołączone krawędziami są wierzchołki przeciwstawne (tj. przystające modulo 3). Jeszcze trzeba poetykietować wierzchołki, powiedzmy, że w H zawsze 1 idzie do mniejszego wierzchołka, zaś w 2 etykietujemy po kolei, od prawego sąsiada 1 do lewego 4. I wykonajmy np. przejście $(2, 1)$ z wierzchołka $(1, 1)$. Wpierw wykonujemy krok (indeksowany dwójką) z wierzchołka $(1, 1)$ do $(1, 3)$. Nie ruszamy pierwszej współrzędnej wierzchołka, wykorzystujemy pierwszy indeks krawędzi. Potem chcemy wykonać krok na G wykorzystując jako etykietę aktualną pozycję w H (czyli tę trójkę) — to nas prowadzi do wierzchołka 3 w G. I teraz jednocześnie musimy podmienić drugą współrzędną na etykietę odwrotną — żeby przejść z 3 do 1 w G idziemy po etykiecie 2, czyli przechodzimy do $(3, 2)$. I teraz krok po grafie H, z drugą współrzędną krawędzi, czyli jedynką, która prowadzi nas z 2 do 1. Czyli doszliśmy z $(1, 1)$ do $(3, 1)$. Pozostali sąsiedzi $(1, 1)$ to $(5, 1)$, $(5, 4)$ i $(3, 4)$ (uwaga, nie zawsze musi być tak, że zbiór sąsiadów ma postać iloczynu kartezjańskiego, tu wynika to ze specyficznej postaci etykietowania na G). A krawędź powrotną dostajemy tak — wpierw musimy pójść z $(3, 1)$ do $(3, 2)$ (czyli pierwsza współrzędna to 2), potem krok grafem G prowadzi nas do $(1, 3)$ (nastąpiła podmiana drugiego wierzchołka), a potem musimy wrócić z 3 do 1 w H, czyli krawędzią powrotną jest $(2, 1)$.

Żeby to sformalizować, wprowadźmy dwie funkcje: $nei(v, i)$, która daje i–tego sąsiada wierzchołka v, oraz $rev(v, i)$, która daje etykietę krawędzi łączącej $nei(v, i)$ z v (czyli $nei(nei(v, i), rev(v, i)) = v$). I wtedy już poprawna definicja wygląda tak: $nei((s, t), (i, j)) = (nei(s, nei(t, i)), nei(rev(s, nei(t, i)), j))$. Żeby zobaczyć, że teraz wyszedł dobry graf, wystarczy zobaczyć, że jeśli $nei((s, t), (i, j)) = (u, v)$, to w drugą stronę poprowadzi nas krawędź $rev(rev(s, nei(t, i)), j), rev(t, i)$.

Teraz oczywiście moc oraz stopień są takie, jak zaplanowaliśmy. Jedyne, co trzeba zrozumieć, to czemu tak właściwie miałby to być dobry ekspander, jeśli grafy wyjściowe były dobrymi ekspanderami. Intuicja jest generalnie z błądzeń losowych — chcemy popatrzeć, dlaczego jeśli mamy jakiś rozkład prawdopodobieństwa na $V(G) \times V(H)$, to krok po naszym $G \otimes H$ przybliża ten rozkład do jednostajnego (to udowodni, że druga wartość własna co do modułu nie jest za duża).

Generalnie w dowodzie będziemy liczyli jak ludzie, po prostu będziemy szacowali drugą normę błędu. Ale w zrozumieniu pomaga — jeśli ktoś go zna — koncept entropii, czyli takiej "miary losowości rozkładu" — rozkład deterministyczny będzie miał entropię zerową, a rozkład jednostajny — największą możliwą. Jeden krok błądzenia losowego po grafie (dowolnym) jest stosunkowo łatwo udowodnić, że nigdy nie zmniejsza entropii, a my chcemy pokazać, że ta entropia zauważalnie rośnie, chyba, że już przed ruchem była nieomal maksymalna. Nie sformalizujemy tu pojęcia "entropia", będziemy tylko traktować ją jako "miarę bliskości do rozkładu jednostajnego" — im większa entropia, tym bliżej jesteśmy rozkładu jednostajnego.

To załóżmy, że nie była maksymalna — czyli nasz rozkład nie był jednostajny. Otóż jeśli rozkład dla pewnego ustalonego $s \in V(G)$ jest daleki od losowego, to losowy krok ma w szczególności losowe i — zatem w pierwszym kroku już zauważalnie wzrośnie entropia, a następnie — w kolejnych krokach — nie będzie spadać. Jeśli dla każdego konkretnego s rozkład na $s \times V(H)$ jest nieomal jednostajny, a cały rozkład nie jest, to znaczy, że rozkład na s–ach jest mocno niejednostajny. Ale wtedy — skoro dla ustalonego s rozkład t jest niemal jednostajny, to jest też niemal jednostajny po pierwszym kroku (entropia nie spada), a zatem krok po grafie G z s jest bardzo bliski krokowi losowemu w grafie G — bo każdą z etykiet wybieramy z prawie

równym prawdopodobieństwem. Wobec tego entropia rozkładu na wierzchołkach H powinna wzrosnąć na tym kroku.

Teraz zauważmy, że drugi krok jest permutacją — tam nie ma wyboru krawędzi, tylko jest po prostu pewna permutacja wierzchołków grafu. Wobec tego ten krok nie może zmienić całościowej entropii. Na tym polega, w istocie, klucz całego pomysłu — nie stać nas na dodawanie entropii za pomocą grafu G — nie chcemy po nim błądzić losowo, bo ma on za duży stopień, i wymaga to od nas nazbyt wielu losowych bitów. Zatem — skoro nastąpił wzrost "entropii rozkładu na G", to musiała jednocześnie zmaleć "entropia rozkładu na H" — czyli jeśli przed krokiem wierzchołki H dla ustalonego s były równo rozłożone, a wierzchołki s — rozłożone nierówno, to po dwóch krokach dla ustalonego s wierzchołki H są rozłożone istotnie mniej równo. Czyli graf G wykorzystaliśmy do "przeniesienia entropii" z rozkładu na H dla ustalonego s gdzie indziej (do rozkładu s–ów), gdzie się nie zmarnuje. Zatem teraz trzeci krok da nam zwiększenie entropii (bo znów błądzimy losowo dla ustalonego s, a zaczynamy od rozkładu mocno niejednostajnego).

To była bajka, a teraz formalnie będziemy formalnie dowodzić własności produktu zygzakowego.

Twierdzenie 2.4. Niech G będzie rozmiaru N i stopnia D, zaś H rozmiaru D i stopnia d. Niech druga wartość własna G co do modułu to αD, druga wartość własna H co do modułu to βd. Wtedy druga wartość własna co do modułu $G \otimes H$ to co najwyżej $(\alpha + \beta + \beta^2)d^2$.

Dowód. Dla dowolnego wektora $a \in \mathbb{R}^{ND}$ (czyli w przestrzeni, na której działa macierz M) będziemy jego współrzędne oznaczali dwuindeksowo, v będzie przebiegało N, zaś k przebiegało D. Weźmy dowolny a spełniający $\sum_v \sum_k a_{v,k} = 0$. Mamy udowodnić, że $|\langle Ma, a\rangle| / \langle a, a\rangle \leqslant (\alpha + \beta + \beta^2)d^2$, gdzie M to macierz sąsiedztwa $G \otimes H$. O wektorze a możemy myśleć jak o "niejednostajnej części" jakiegoś rozkładu prawdopodobieństwa (albo o odchyleniu od rozkładu jednostajnego).

Przypomnijmy sobie, że rozważaliśmy w heurezie dwa przypadki — albo a jest niejednostajny na którejś grupie D wierzchołków (i wtedy pierwszy krok powinien mu pomóc), albo jest jednostajny na każdej grupie, ale wtedy masy grup są nierówne. Wprowadźmy więc dla $v \in N$ oznaczenie a_v na wektor $(a_{v,1}, a_{v,2}, \ldots, a_{v,D})$ (to jest odchylenie na pojedynczej grupie) oraz przekształcenie liniowe $Ca : \mathbb{R}^{ND} \to \mathbb{R}^N$, zadane przez $(Ca)_v = \sum_{k=1}^{D} a_{v,k}$ (tu uśredniamy wewnątrz grup, i dostajemy rozkład na wierzchołkach z N). Każdy a_v rozkłada się na część zależną od $\mathbf{1}_D$ (ozn. $a_v^{||}$) oraz część prostopadłą (ozn. $a_v^{\perp}$). Jako, że a jest sumą odpowiednich a_v (formalnie — $a_v \otimes e_v$, czyli wektorów b, dla których $b_{w,k} = 0$ dla $w \neq v$ i $b_{v,k} = a_{v,k}$), to otrzymujemy rozkład $a = \sum_v a_v \otimes e_v = \sum_v a_v^{||} \otimes e_v + \sum_v a_v^{\perp} \otimes e_v$, pierwszą część oznaczmy $a^{||}$, drugą zaś $a^{\perp}$.

Zauważmy, że każdy $a_v^{\perp}$ jest prostopadły do $\mathbf{1}_D$, zatem $a^{\perp}$ jest prostopadły do $\mathbf{1}_{ND}$, skąd $a^{||}$ też jest prostopadły do $\mathbf{1}_{ND}$ (bo a jest), a zatem Ca jest prostopadły do $\mathbf{1}_N$.

Spróbujmy, choć częściowo, rozpisać macierz M jako produkt macierzy odpowiadających poszczególnym krokom. Niech A będzie macierzą G, zaś B — macierzą H. Wtedy pierwszy oraz trzeci krok odpowiadają macierzy $\tilde{B} = I \otimes B$ — nie ruszamy pierwszej współrzędnej, drugą robimy połączenia wg B. Macierz $\tilde{A}$ środkowego kroku jest macierzą permutacji (to wiemy, tam każdy wierzchołek łączy się z dokładnie jednym wierzchołkiem), i jest symetryczna (bo dostajemy graf nieskierowany). Więcej powiemy o niej za chwilę. Oczywiście $\tilde{B}$ też jest symetryczna.

Badamy $\langle Ma, a\rangle = \left\langle \tilde{B}\tilde{A}\tilde{B}a, a\right\rangle = \left\langle \tilde{A}\tilde{B}a, \tilde{B}a\right\rangle$ z symetrii $\tilde{B}$. Zauważmy, że $\tilde{B}a^{||} = da^{||}$ (bo $a^{||}$ na każdej grupie o ustalonym v jest stałe, czyli przy mnożeniu przez B mnoży się przez stopień B). Wobec tego $\langle Ma, a\rangle = \left\langle \tilde{A}\tilde{B}(a^{||} + a^{\perp}), \tilde{B}(a^{||} + a^{\perp})\right\rangle = \left\langle d\tilde{A}a^{||} + \tilde{A}\tilde{B}a^{\perp}, da^{||} + \tilde{B}a^{\perp}\right\rangle$. Szacujemy to z góry przez $d^2\left\langle \tilde{A}a^{||}, a^{||}\right\rangle + d\|\tilde{A}a^{||}\|\|\tilde{B}a^{\perp}\| + d\|\tilde{A}\tilde{B}a^{\perp}\|\|a^{||}\| + \|\tilde{A}\tilde{B}a^{\perp}\|\|\tilde{B}a^{\perp}\|$. Zauważmy

jeszcze, że $\tilde{A}$ jest macierzą permutacji, a więc nie zmienia normy wektorów, stąd możemy ją usunąć tam, gdzie szacujemy przez normę, i dostajemy $d^2|\left\langle \tilde{A}a^{||}, a^{||}\right\rangle|+2d\|a^{||}\|\|\tilde{B}a^{\perp}\|+\|\tilde{B}a^{\perp}\|^2$.

Teraz już widać, skąd się będą brały poszczególne człony w naszym oszacowaniu.

Lemat 2.5. $\|\tilde{B}a^{\perp}\| \leqslant \beta d\|a^{\perp}\|$.

Dowód. $a^{\perp} = \sum_v (a_v^{\perp}) \otimes e_v$. $\tilde{B}a^{\perp} = \sum_v (B \otimes I)(a_v^{\perp} \otimes e_v) = \sum_v (Ba_v^{\perp}) \otimes e_v$. Wszystkie składniki sumy są prostopadłe, dla każdego mamy $\|Ba_v^{\perp}\| \leqslant \beta d\|a_v^{\perp}\|$, co kończy dowód. □

Lemat 2.6. $|\left\langle \tilde{A}a^{||}, a^{||}\right\rangle| \leqslant \alpha \left\langle a^{||}, a^{||}\right\rangle$.

Dowód. Zauważmy, że fortunnie nie musimy rozumieć, jak, tak w ogóle, działa $\tilde{A}$. Wystarczy zauważyć, że skoro dla każdego v wektor $a_v^{||}$ jest rozłożony jednostajnie, to $\left\langle a^{||}, \tilde{A}a^{||}\right\rangle = \left\langle Ca^{||}, C\tilde{A}a^{||}\right\rangle /D$, bo $(a^{||})_{v,k} = (Ca^{||})_v/D$, zatem

$$\sum_v \sum_k (a^{||})_{v,k}(\tilde{A}a^{||})_{v,k} = \sum_v (Ca^{||})_v \sum_k (\tilde{A}a^{||})_{v,k}/D = \sum_v (Ca^{||})_v (C\tilde{A}a^{||})_v/D = \left\langle Ca^{||}, C\tilde{A}a^{||}\right\rangle /D.$$

Dalej zauważmy, że $DC\tilde{A}a^{||} = ACa^{||}$, bo wyraz v obydwu stron jest równy $\sum_{w:wv\in E(G)} (Ca^{||})_w$. Wobec tego $\left\langle a^{||}, \tilde{A}a^{||}\right\rangle = \left\langle Ca^{||}, ACa^{||}\right\rangle /D^2$. Fortunnie już udowodniliśmy, że $Ca^{||}$ jest prostopadłe do wektora N jedynek, skąd ten iloczyn jest nie większy niż $\alpha D \left\langle Ca^{||}, Ca^{||}\right\rangle /D^2 = \alpha \left\langle a^{||}, a^{||}\right\rangle$. □

Aby zakończyć dowód, weźmy $p = \|a^{||}\|/\|a\|$, $q = \|a^{\perp}\|/\|a\|$, wtedy $p^2 + q^2 = 1$, i mamy $|\langle Ma, a\rangle|/\langle a, a\rangle \leqslant d^2(\alpha p^2 + 2\beta pq + \beta^2 q^2) \leqslant d^2(\alpha + \beta + \beta^2)$ (szacujemy trywialnie p i q przez 1, a pq przez 1/2). □

Warto jeszcze zauważyć, że to ograniczenie górne, które napisaliśmy, może wyjść ponad d^2 dla stosunkowo dużych α i β. W takim wypadku podamy lepsze ograniczenie:

Wniosek 2.7. Przy założeniach jak wyżej, druga co do modułu wartość własna M jest nie większa niż $d^2 f(\alpha, \beta)$, gdzie $f(\alpha, \beta)$ jest ostro mniejsze od d^2 dla $\alpha, \beta < 1$.

Meritum tego ograniczenia jest takie, że nie zależy ono od d, D ani N, a tylko od jakości ekspansji zygzakowanych grafów — fakt, że zygzak dwóch ekspanderów jest ekspanderem (tj. jest spójny i nie dwudzielny) nie jest trudny, ale tu możemy jakość ekspansji oszacować niezależnie od parametrów wejściowych.

Dowód. Załóżmy $\|a\| = 1$. Jeśli $\|a^{\perp}\|$ jest małe, powiedzmy $q \leqslant \frac{1-\alpha}{3}$, to szacowanie, które mamy powyżej jest dobre: szacujemy przez $d^2(\alpha + 2\frac{1-\alpha}{3} + \frac{(1-\alpha)^2}{9}) \leqslant d^2\frac{\alpha+8}{9}$, gdzie α^2 i β oszacowaliśmy z góry przez 1. Czyli w tym wypadku się udało.

Jeśli q jest duże, to — wedle rozwijanych intuicji — ten środkowy krok powinien generalnie być zbędny. Szacujemy więc $\langle Ma, a\rangle = \left\langle \tilde{A}(da^{||} + \tilde{B}a^{\perp}), da^{||} + \tilde{B}a^{\perp}\right\rangle \leqslant \|da^{||} + \tilde{B}a^{\perp}\|^2$. Zauważmy, że $\left\langle a^{||}, \tilde{B}a^{\perp}\right\rangle = d\left\langle a^{||}, a^{\perp}\right\rangle = 0$, czyli dwa wektory pod normą są prostopadłe, zatem całość $= d^2\|a^{||}\|^2 + \|\tilde{B}a^{\perp}\|^2 = d^2(1 - \|a^{\perp}\|) + \|\tilde{B}a^{\perp}\|^2$. To drugie już liczyliśmy że szacuje się przez $d\beta\|a^{\perp}\|$, więc mamy w sumie $d^2(1 - \frac{(1-\alpha)^2}{9}(1 - \beta^2))$. □

Teraz jeszcze uwaga — umiemy dostać całkiem przyzwoitą ekspansję wierzchołkową (konkretnie 21/29), ale niestety stopień grafu, który otrzymamy, jest dość duży (należy się liczyć z tym, że koło 1000). Można sobie z tym poradzić, redukując stopień (oczywiście kosztem pogorszenia ekspansji) poprzez, choćby, zzygzakowanie na koniec z cyklem nieparzystym długości równej stopniowi grafu (jeśli stopień grafu wyszedł parzysty, to możemy w każdym wierzchołku dołożyć pętlę, to nie pogarsza ekspansji).

3 Wykład szósty — zastosowania ekspanderów — redukcja losowości oraz $L = SL$

3.1 $L = SL$

Problem który rozważamy jest następujący: mamy dany nieskierowany graf G o n wierzchołkach i dwa wierzchołki s i t. Czy da się dojść z s do t? Oczywiście, dowolny algorytm przeszukiwania grafu rozwiązuje ten problem. Ale potrzebuje $O(n)$ pamięci. My zaś chcemy to zrobić w $O(\log n)$ pamięci. Czyli, de facto, możemy trzymać stałą liczbę indeksów o zakresie wielomianowym w n.

Klasa problemów rozwiązywalnych deterministycznie w logarytmicznej pamięci to L. W klasie NL mamy też logarytmiczną pamięć, ale niedeterministyczną maszynę Turinga. Oczywiście więc nasz problem jest w NL. Co więcej, można pokazać, że nasz problem, ale na grafach skierowanych, jest NL–zupełny. To, gdzie leży nasz problem pomiędzy L a NL było problemem otwartym przez długo, do tego stopnia, że stworzono specjalną klasę SL: problemy redukowalne do naszego problemu. W 2005 Omar Reingold pokazał, że $SL = L$, używając zygzaka. Zajmiemy się teraz pokazaniem jego algorytmu.

Sprecyzujmy sytuację: mamy maszynę Turinga, na taśmie tylko do odczytu mamy dany nasz graf i startowe wierzchołki, oraz mamy $O(\log n)$–bitową taśmę do bazgrania. Zwróćmy uwagę, że możemy dość swobodnie odczytywać nasz graf, bo wszelkie indeksy i wskaźniki na taśmę wejściową mają właśnie $O(\log n)$ bitów.

Wpierw prosta obserwacja: możemy założyć, że G jest 3–regularny. Zastępujemy każdy wierzchołek v w G cyklem złożonym z $\deg_G(v)$ wierzchołków; każdy wierzchołek na cyklu ma jedną krawędź incydentną spoza cyklu. Tym samym, w nowym grafie wierzchołki są etykietowane (v, i), gdzie $v \in V(G)$ i $1 \leqslant i \leqslant \deg_G(v)$; krawędzie incydentne z (v, i) prowadzą do $(v, i \pm 1)$ oraz (u, j), jeśli i–ta krawędź z v prowadziła do u, a j–ta krawędź z u prowadziła do v. Widać, że można to bez problemu symulować w L.

3.1.1 Algorytm randomizowany

Dość prostym pomysłem jest: zaczynamy z s, błądzimy losowo po grafie, i po pewnej liczbie kroków, jeśli nie trafimy do t, to mówimy NIE. Ile kroków potrzeba? Mając narzędzia z pierwszego wykładu z ekspanderów szybko pokażemy, że robiąc $O(n^2)$ wycieczek po $O(n^3)$ kroków każda, będzie dobrze. Dla ułatwienia technicznej analizy, dostawmy w każdym wierzchołku G pętelkę.

To jest oczywiste i było na ćwiczeniach:

Lemat 3.1. Jeśli G jest spójny i w każdym wierzchołku ma pętelkę, to $h^V(G) \geqslant 2/n$.

Korzystając z twierdzenia 1.17 otrzymujemy:

Lemat 3.2. Jeśli G jest spójny, w każdym wierzchołku ma pętelkę i jest d–regularny, to

$$\Delta'(G) \geqslant d - \sqrt{d^2 - \frac{4}{2d^2n^2}} \geqslant \frac{1}{d^3n^2}.$$

Przypomnijmy sobie analizę błądzenia losowego. Tutaj zaczynamy od rozkładu

$$p^{(0)} = e_s = \sum_{i=1}^{n} a_i^{(0)} v_i.$$

W związku z tym ($d = 3$) po około $k := 3^3 n^3$ krokach współczynniki przy v_i dla $i \geqslant 2$ będą rzędu:

$$a_i^{(k)} \leqslant (1 - \Delta'(G))^k \sim e^{-n}.$$

Czyli rozkład będzie prawie jednostajny: każda taka $O(n^3)$–krokowa wycieczka de facto kończy w losowym wierzchołku, do którego da się dojść z s. Robiąc n^2 takich wycieczek, prawdopodobieństwo, że któryś wierzchołek ze spójnej składowej z s pominiemy wynosi

$$\left(1 - \frac{1}{n}\right)^{n^2} \sim e^{-n}.$$

Czyli znów malutko.

Pytanie więc brzmi: jak zderandomizować ten pomysł?

3.1.2 Algorytm z pamięcią $O(\log^2 n)$

Przypomnijmy w G^2 zbiór wierzchołków jest taki sam, a każda krawędź odpowiada ścieżce długości 2. Niech $k := \lceil \log n \rceil$. Chcemy zrobić G^{2^k} i przeiterować wszystkich sąsiadów s. Ich będzie 2^k, czyli wielomianowo w n.

Fajny pomysł, ale jak to zrobić. Sprecyzujmy przeiterowanie: w grafie d–regularnym G definiujemy operację $\mathrm{Rot}_G : V(G) \times \{1, 2, \ldots, d\} \to V(G) \times \{1, 2, \ldots, d\}$. $\mathrm{Rot}_G(v, i) = (u, j)$ jeśli uv jest i-tą krawędzią wychodzącą z i oraz j-tą krawędzią wychodzącą z u. Ile pamięci na taśmie do bazgrania potrzebujemy, by policzyć $\mathrm{Rot}_{G^2}(v, (i_1, i_2))$? Robimy $\mathrm{Rot}_G(v, i_1) = (u', j_2)$ i następnie $\mathrm{Rot}_G(u', i_2) = (v, j_1)$ i zwracamy $(v, (j_1, j_2))$. Potrzebujemy umieć liczyć Rot_G i dodatkowo trzymać indeksy wierzchołków ($O(\log n)$) i indeksy krawędzi ($O(\log \deg(G)) = O(\log n)$). Czyli, by policzyć $\mathrm{Rot}_{G^{2^k}}$, potrzebujemy k razy się zagłębić rekurencyjnie. Otrzymyjemy złożoność pamięciową $O(\log^2 n)$.

Prosto możemy zaoszczędzić pamięć $O(\log n)$ w każdym kroku rekurencji, zużytą na indeksy wierzchołków. Załóżmy, że chcemy zrobić Rot takie, które na taśmie do bazgrania dostaje (v, i) i chce w tym samym miejscu nadpisać wynik. Zauważmy, że Rot_{G^2} da się tak zrobić, mając tak działające Rot_G: w trakcie działania, tylko przepisujemy różne indeksy i_1, i_2, j_1, j_2. Czyli potrzebujemy tylko jednego miejsca na indeks wierzchołka, wyliczając $\mathrm{Rot}_{G^{2^k}}(s, i)$.

Nie widać natomiast, jak zaoszczędzić $O(\log \deg(G^{2^i}))$ na każdym kroku: stopień nam strasznie eksploduje w tej konstrukcji.

3.1.3 Algorytm Reingolda

Pomysł jest następujący. Robimy tak jak poprzedni algorytm, ale co jakiś czas graf G zygzakujemy z stałym grafem H i w ten sposób redukujemy stopień G spowrotem do $(\deg(H))^2$. Zygzakowanie zapewni nam, że graf cały czas będzie miał dobrą (i rosnącą) ekspansję, i po niewielkiej (logarytmicznej, bo tylko na taki stos mamy miejsce) liczbie kroków będzie blisko z s do t.

A teraz precyzyjnie. Ustalmy stałą d i zróbmy z G graf d^{16}–regularny. Po prostu robimy go wpierw 3–regularnym, a następnie dodajemy $d^{16} - 3$ pętelek w każdym wierzchołku. To da się zasymulować w logarytmicznej pamięci. Co więcej, te pętelki dają nam $\Delta'(G) = \Omega(n^{-2})$ z lematu 3.2.

Weźmy ustalony graf H o d^{16} wierzchołkach, d–regularny i o $\Delta'(H) \geqslant d/2$ — czyli bardzo dobry ekspander. Na ćwiczeniach pokazaliśmy lub pokażemy, że takie istnieją; tutaj H jest whardkodowany w algorytm i o stałej wielkości.

Ustalmy L jako najmniejszą liczbę naturalną taką, że

$$\left(1 - \frac{1}{d^{48}n^2}\right)^{2^L} \leqslant \frac{1}{2}.$$

To jest spełnione, jak $2^L = O(d^{48}n^2)$, czyli $L = O(\log d + \log n) = O(\log n)$.

Teraz L razy iterujemy operację $G_{i+1} = (G_i \otimes H)^8$, $G_0 = G$. Zauważmy, że z własności zygzaka i podnoszenia grafu do potęgi otrzymujemy od razu, że każdy G_i jest d^{16}-regularnym grafem. Teraz pora pokazać, że jest dobrym ekspanderem.

Lemat 3.3.

$$\Delta'(G_L)/d^{16} \geqslant 1/2.$$

By tego dowieść, potrzebujemy niestety trochę lepszego oszacowania na $\Delta'(G \otimes H)$, niż udowodniliśmy na poprzenim wykładzie. Nie będziemy go tu dowodzić, bo jest dość syfny. Przypomnijmy, że oznaczaliśmy $\beta = 1 - \Delta'(H)/d$.

Twierdzenie 3.4.

$$\Delta'(G \otimes H)/d^2 \geqslant (1 - \beta^2)/2 \cdot \Delta'(G)/d.$$

Prostym wnioskiem jest teraz:

Lemat 3.5. Jeśli $\Delta'(H) \geqslant d/2$, tj., $\beta \leqslant 1/2$, to $\Delta'(G \otimes H)/d^2 \geqslant 3/8 \cdot \Delta'(G)/d$.

Oczywistym wnioskiem z tego, jak się potęguje macierze, jest że:

Lemat 3.6.

$$\Delta'(G^r)/d^r = 1 - (1 - \Delta'(G)/d)^r.$$

Pora przystąpić do dowodu lematu 3.3.

Dowód lematu 3.3. Wpierw zauważmy, że jeśli $\Delta'(G_{i-1})/d^{16} \geqslant 1/2$, to

$$\Delta'(G_i)/d^{16} \geqslant 1 - \left(1 - \Delta'(G_{i-1} \otimes H)/d^2\right)^8 \geqslant 1 - \left(1 - \frac{3}{16}\right)^8 > \frac{1}{2}.$$

Czyli, jeśli któryś G_i będzie miał $\Delta'(G_i)/d^{16} \geqslant 1/2$, to wszystkie późniejsze tym bardziej.

Teraz chcemy oszacować, jak szybko $\Delta'(G_i)/d^{16}$ zbiega do $1/2$. Łatwo sprawdzić, np. różniczkując, że

$$\left(1 - \frac{3}{8}x\right)^4 \leqslant 1 - x \qquad \text{dla } 0 \leqslant x \leqslant \frac{1}{2}.$$

Mamy więc

$$1 - \Delta'(G_{i+1})/d^{16} = (1 - \Delta'(G_i \otimes H)/d^2)^8 \leqslant \left(1 - \frac{3}{8}\Delta'(G_i)/d^{16}\right)^8 \leqslant (1 - \Delta'(G_i)/d^{16})^2.$$

Wobec tego

$$1 - \Delta'(G_L)/d^{16} \leqslant (1 - \Delta'(G)/d^{16})^{2^L}.$$

Z lematu 3.2 tego, jakie L przyjęliśmy, wynika, że $1 - \Delta'(G_L)/d^{16} < 1/2$, co kończy dowód. □

Co oznacza, że $\Delta'(G_L)/d^{16} > 1/2$? Z twierdzenia 1.17 wynika, że daje to $h^V(G_L) > c$ dla pewnej stałej c. Niech $B_x(s) = \{v \in V(G_L) : d_{G_L}(s,v) \leqslant x\}$. Z definicji $h^V(G_L)$ wynika, że $|B_{x+1}(s)| \geqslant (1+c)B_x(s)$ jeśli $|B_x(s)| < n/2$. Czyli dla $x_0 := \log_{1+c}(n/2) = O(\log n)$ mamy $|B_{x_0}(s)| \geqslant n/2$. Analogicznie dla t; wobec tego $d_{G_L}(s,t) \leqslant 2x_0$ o ile leżą w tej samej spójnej składowej.

G_L jest d^{16}–regularny. Puszczamy więc brucika, przeglądającego wszystkie ścieżki długości $2x_0$ wychodzące z s i patrzący, czy nie dojdziemy do t. Powyższa analiza pokazuje, że algorytm ten jest poprawny. Trzeba chwilę posztuczkować, by pokazać, że rzeczywiście da się go zrealizować używając tylko $O(\log n)$ pamięci.

Stos brucika ma wielkość $O(\log n)$, więc mamy $O(1)$ pamięci na każdą ramkę stosu rekurencji brucika. Możemy więc tylko zapamiętać indeks krawędzi wychodzącej w wierzchołku na stosie; nie możemy zapamiętać indeksu wierzchołka. Potrzebujemy więc, by chodzić skutecznie po G_L, implementacji Rot_{G_L}, zużywającej $O(L) = O(\log n)$ pamięci.

Analizując algorytm z pamięcią $O(\log^2 n)$ pokazaliśmy, że da się robić Rot_{G^2} mając pamięc $O(\log \deg(G))$. Tak samo można zrealizować $\mathrm{Rot}_{G\otimes H}$, a nawet prościej. Zauważmy, że $|H| = O(1)$, więc całe chodzenie po H możemy zrealizować w stanach maszyny Turinga.

Wobec tego mamy Rot_{G_L} wyliczające się $O(L)$ pamięci. Trzymamy na stosie brucika tylko indeksy krawędzi, którymi poszliśmy, oraz powrotne indeksy tychże krawędzi. Indeks wierzchołka pamiętamy tylko jeden: ten, w którym aktualnie jesteśmy. Aby pójśc dalej, wyliczamy Rot_{G_L} i odkładamy numer krawędzi powrotnej na stosie.

3.2 Analiza redukcja losowości z wykładu czwartego

Przypomnienie sekcji 1.2.1. Mamy algorytm $A(x,y)$, co bierze y — ciąg bitów losowych długości $m = m(|x|)$ i rozstrzyga, czy $x \in L$. Jeśli $n \notin L$, to A zawsze zwraca NIE, zaś w przeciwnym przypadku może zwrócić NIE z prawdopodobieństwem nie większym niż $p_A < 1/4$. Powtarzając A k razy, mamy prawdopodobieństwo błędu co najwyżej p_A^k, lecz używamy km bitów losowych. Chcielibyśmy tak zredukować liczbę bitów losowych.

1. Weźmy G — graf o 2^m wierzchołkach (etykietowanych ciągami m–bitowymi), d–regularny, i o dobrej ekspansji.
2. Wylosujmy y_0 — wierzchołek G i puśćmy $A(x,y_0)$.
3. Następnie k razy wylosujmy y_{i+1} — sąsiada y_i — i puśćmy $A(x,y_{i+1})$.

Używamy tu tylko $m + k\log d$ bitów losowych. Pytanie jest, jaka jest szansa błędu w tym algorytmie.

Ustalmy $x \in L$ i niech $Z = \{y : A(x,y) = 0\}$, czyli zbiór tych y, że algorytm A się pomyli dla wejścia x. Oczywiście $|Z| \leqslant p_A n < n/4$. Niech $Z(t)$ oznacza zdarzenie polegające na tym, że $y_0, y_1, \ldots, y_t \in Z$. Nasze prawdopodobieństwo błędu to prawdopodobieństwo zdarzenia $Z(k)$. Niech λ_2 oznacza drugą wartość własną $M(G)$. Poniższe twierdzenie zawiera kluczowe szacowania, gdzie $\lambda = \max(|\lambda_n|, |\lambda_2|)$.

Twierdzenie 3.7.

$$\mathbb{P}(Z(t)) \leqslant \left(\frac{\lambda}{d} + \frac{|Z|}{n}\right)^t.$$

Zanim udowodnimy to twierdzenie, pokażmy, że z niego od razu wynika teza. Mamy

$$\mathbb{P}(Z(k)) \leqslant \left(\frac{\lambda}{d} + p_A\right)^k.$$

Czyli prawdopodobieństwo błędu maleje też wykładniczo. Podstawa jest trochę gorsza niż p_A, ale jeśli G jest dobrym ekspanderem — czyli λ/d jest dość małe — a przy tym d też jest małe — to i tak zaoszczędzimy sporo bitów losowych.

Dowód twierdzenia 3.7. Bawimy się, jak zawsze, algebrą liniową nad przestrzenią rozpiętą przez $\{e_v : v \in V(G)\}$. Oznaczmy przez P operator rzutowania na podprzestrzeń rozpiętą przez $\{e_v : v \in Z\}$. Oznaczmy też $u = \frac{1}{\sqrt{n}} v_1$ — rozkład jednostajny na $V(G)$. Zauważmy, że

$$\mathbb{P}(Z(0)) = \|Pu\|_1.$$

Teraz już tylko krok do zauważenia, że

$$\mathbb{P}(Z(t)) = \|(P\bar{M}P)^t Pu\|_1,$$

gdzie $\bar{M} = M(G)/d$ jest macierzą przejścia błądzenia losowego. Zauważmy, że

$$\mathbb{P}(Z(t)) \leqslant \sqrt{n}\|(P\bar{M}P)^t Pu\|_2 \leqslant \sqrt{n}\|P\bar{M}P\|^t\|Pu\|_2 = \sqrt{n}\|P\bar{M}P\|^t \frac{\sqrt{|B|}}{n} \leqslant \|P\bar{M}P\|^t.$$

Pozostaje więc wykazać, że

$$\|P\bar{M}P\| \leqslant \frac{\lambda}{d} + \frac{|Z|}{n},$$

gdzie $\|P\bar{M}P\|$ to oczywiście norma operatora w ℓ_2, $\|P\bar{M}P\| = \sup_{\|v\|_2=1} \|P\bar{M}Pv\|_2$.

Szukamy więc tak naprawdę największego modułu wartości własnej $P\bar{M}P$, czyli chcemy zmaksymalizować $|v^T P\bar{M}Pv|$ po $\|v\|_2 = 1$. Z zwartości, weźmy maksymalizujące v i niech

$$Pv = \sum_{i=1}^{n} a_i v_i.$$

Oczywiście $\sum_{i=1}^n a_i^2 = \|Pv\|_2^2 \leqslant 1$. Oznaczmy $v' = \sum_{i=2}^n a_i v_i = Pv - a_1 v_1$. Mamy więc ($P^T = P$):

$$v^T P\bar{M}Pv = (Pv)^T\bar{M}(Pv) = (a_1v_1 + v')^T\bar{M}(a_1v_2 + v') = a_1^2 v_1^T\bar{M}v_1 + v'^T\bar{M}v' = a_1^2 + \frac{\lambda}{d}.$$

Pozostaje oszacować a_1:

$$a_1 = \langle Pv, v_1\rangle = \langle Pv, Pv_1\rangle \leqslant \|Pv\|_2\|Pv_1\|_2 \leqslant \sqrt{\frac{|Z|}{|n|}}.$$

□

4 Wykład siódmy — PCP

4.1 Remanent — bezwzględna przerwa spektralna a ekspansja wierzchołkowa

Wpierw remanent z poprzednich wykładów. Zaczniemy od uzupełnienia kompletu nierówności wiążących przerwy spektralne z ekspansjami. Marcin zrobił już połączenie przerwy względnej z ekspansją krawędziową, przypomnijmy:

$$\Delta(G)/2 \leqslant h^E(G) \leqslant \sqrt{2d\Delta(G)}.$$

Marcin pokazał też prosty fakcik, że $h_V(G^2) \geqslant h_V(G)$ — przypomnijmy, że jest on oczywisty. Teraz udowodnimy brakujący komplet nierówności:

Twierdzenie 4.1. Niech G będzie grafem d–regularnym. Wtedy

$$\frac{d^2-(d-\Delta'(G))^2}{d^2+(d-\Delta'(G))^2} \leqslant h^V(G) \leqslant d\sqrt{2}\sqrt{d^2-(d-\Delta'(G))^2}.$$

Dowód. Będziemy rozważać związki między $\|v\|$ a ekspansją. Tu nie trzeba robić tej sztuczki z odejmowaniem, bo interesuje nas druga co do modułu wartość własna, czyli po prostu ekstremalizujemy $\|Av\|/\|v\|$ po $\sum v_i = 0$. Zauważmy, że $\|v\|^2 \geqslant 1/|\mathrm{supp}v|$ dla $\|v\|_1 = 1$ (to jest nierówność CBS: $1 = \|v\|_1 \leqslant \sqrt{|\mathrm{supp}v|}\|v\|$). No to weźmy dowolny wektor $v = \mathbf{1}_S/|S|$. Przepuszczamy go przez macierz A, dostajemy Av, i $\|Av\|^2 \geqslant d^2/|\mathrm{supp}Av| = d^2/|\Gamma(S)|$, zaś $\|v\|^2 = 1/|S|$. Teraz każdy z tych wektorów po odjęciu n będzie prostopadły do n, czyli $1/|S| = \|v\|^2 = \|n\|^2 + \|v'\|^2 = 1/N + \|v'\|$, oraz $d^2/|\Gamma(S)| \leqslant \|Av\|^2 = \|An + Av'\|^2 = \|dn + Av'\|^2 = d^2\|n\|^2 + \|Av'\|^2 \leqslant d^2/N + \lambda^2\|v'\|^2$. Dzielimy drugą równość przez d^2 i wstawiamy $\|v'\|$ z pierwszej, dostajemy $1/|\Gamma(S)| \leqslant 1/N + (\lambda/d)^2(1/|S| - 1/N)$. Teraz $N \geqslant 2|S|$, i $1/N$ jest po prawej z dodatnim współczynnikiem, czyli $1/|\Gamma(S)| \leqslant (\lambda/d)^2/|S| + (1-(\lambda/d)^2)/2|S|$, co daje $|\Gamma(S)| \geqslant |S|\frac{2}{(\lambda/d)^2+1}$, czyli $h^V(G) \geqslant \frac{d^2-(d-\Delta')^2}{d^2+(d-\Delta')^2}$.

To teraz jeśli mamy ekspansję wierzchołkową h dla G, to mamy wierzchołkową h dla G^2 z faktu powyżej, to mamy krawędziową h dla kwadratu G^2 (bo wierzchołkowa jest nie większa niż krawędziowa) to mamy względną przerwę dla kwadratu (z twierdzenia poprzedniego), to mamy bezwzględną przerwę dla grafu pierwotnego (bo wartości własne kwadratu macierzy to kwadraty wartości własnych macierzy). I już, dostajemy, mam wrażenie, że $\Delta'_G \geqslant d - \sqrt{d^2 - \frac{(h^V)^2}{2d^2}}$. To nie jest najlepszy możliwy wynik, bo jak mamy ekspansję wierzchołkową, to da się wydusić lepszą ekspansję spektralną niż tylko przechodząc przez krawędzie, ale jest akceptowalny, bo dowodzi, że jak jest ekspansja wierzchołkowa niezależna od N, to jest i spektralna niezależna od N. □

To powyżej z góry odpowiada szacowaniu $\Delta'(G) \geqslant d - \sqrt{d^2 - \frac{h^V(G)^2}{2d^2}}$. Da się wyciągnąć lepiej:

Twierdzenie 4.2. $\Delta'(G) \geqslant d - \sqrt{d^2 - \frac{h^V(G)^2}{8+4h^V(G)^2}}$.

4.2 Twierdzenie PCP — wstęp

Będziemy dowodzić twierdzenia zwanego PCP, czyli twierdzenia o trudności aproksymacji pewnych problemów NP–zupełnych. Wpierw zdefiniujemy problem, który będziemy rozważać:

Definicja 4.3. Niech V będzie pewnym skończonym zbiorem, Σ — skończonym alfabetem stałego rozmiaru, zaś q — ustaloną liczbą. Niech $U \subset V$, $|U| = q$, wtedy ograniczeniem na U nazwiemy podzbiór $c_U \subset \Sigma^q$. Rodziną ograniczeń C na V nazwiemy rodzinę ograniczeń na pewnej rodzinie podzbiorów V. Mówimy, że wartościowanie $\phi : V \to \Sigma$ jest zgodne z ograniczeniem c_U na U, jeśli ϕ obcięte do U należy do c_U (czyli c_U zadaje zbiór dopuszczalnych wartościowań na U). I naturalnie wartościowanie ϕ jest zgodne z rodziną ograniczeń, jeśli jest zgodne z każdym ograniczeniem tej rodziny.

W problemie spełniania ograniczeń pytamy o to, czy dla danego V, q, Σ oraz C istnieje wartościowanie zgodne z C.

Będziemy myśleć o q i Σ jako o stałych, zaś zmiennymi będą rozmiar V i rozmiar C. Zwróćmy uwagę, że pojedyncze c_U ma najwyżej $|\Sigma|^q$ elementów, czyli stałą liczbę, a różnych podzbiorów V jest $\binom{|V|}{q}$, zatem przy stałych $|\Sigma|$ i q rozmiar C jest wielomianowy względem rozmiaru V.

Zauważmy też, że przykładowo dla $q = 2$ i $|\Sigma| = 3$ problem spełniania ograniczeń jest trudniejszy niż 3–kolorowanie, zaś dla $q = 3$ i $|\Sigma| = 2$ problem spełniania ograniczeń jest

trudniejszy niż 3–CNF-SAT; z drugiej strony dla $q = |\Sigma| = 2$ ten problem rozwiązuje się wielomianowo przez 2–SAT, co daje nam kompletną listę parametrów, dla których ten problem jest NP–zupełny (to, że jest NP, jest trywialne).

Definicja 4.4. Niech $G = (V, \Sigma, q, C)$ będzie instancją problemu spełniania ograniczeń. Niech ϕ będzie wartościowaniem, wtedy przez $UNSAT(\phi)$ oznaczymy ułamek ograniczeń, które są niespełnione, czyli $\mathbb{P}_{c_U \in C}(\phi|_U \notin c_U)$. Przez $UNSAT(G)$ oznaczymy $\inf_\phi UNSAT(\phi)$.

W optymalizacyjnym problemie spełniania ograniczeń pytamy o wartość $UNSAT(G)$.

Naszym celem będzie udowodnienie twierdzenia PCP:

Twierdzenie 4.5 (PCP, wersja aproksymacyjna). Rozważmy problem, który na wejściu dostaje $G = (V, \Sigma, q, C)$, przy czym wiemy, że $UNSAT(G) = 0$ lub $UNSAT(G) \geqslant 1/2$. Istnieją takie q i $|\Sigma$, dla których rozstrzygnięcie pomiędzy tymi dwoma możliwościami jest NP–trudne.

Dla przypadku $q = 2$, który będziemy przez prawie całą resztę wykładu rozważać, mamy naturalny graf (V, E), gdzie E to zbiór tych par wierzchołków V, dla których zostały zdefiniowane pewne ograniczenia (tj. $c_E \in C$).

Ten wynik ma znaczenie w kilku różnych kierunkach, które być może omówimy na przyszłym wykładzie, w szczególności daje się z niego wydedukować nieaproksymowalność szeregu innych problemów (np. kliki, set–covera, max-3-sata). Teraz jednak skupmy się na dowodzie. Przez $|G|$ oznaczymy $|V| + |E|$. Jeszcze uwaga o zrozumieniu — te dwa dowody ($L = SL$ i PCP) prezentujemy jako przykład zastosowania ekspanderowego sposobu podejścia do problemów. Jest w nich sporo części wspólnych i podobnych pomysłów, na które zdecydowanie warto zwrócić uwagę, zrozumienie jednego z tych dowodów zapewne pomoże w zrozumieniu drugiego.

Zaczniemy od trywialnego faktu (nieco podobnego do tego, od czego zaczynał się dowód $L = SL$):

Fakt 4.6. Jeśli $UNSAT(G) \neq 0$, to $UNSAT(G) \geqslant 1/|G|$ — jeśli nie da się spełnić wszystkich warunków, to zawsze przynajmniej jeden z $|E|$ warunków będzie niespełniony.

Naszym głównym narzędziem będzie zatem twierdzenie o amplifikacji:

Twierdzenie 4.7 (o amplifikacji). Istnieje taki alfabet Σ_0, że dla każdego Σ istnieją stałe $C > 0$ i $1 > \alpha > 0$, oraz algorytm. Algorytm bierze na wejściu graf ograniczeń $G = (V, E, \Sigma, C)$, i zwraca graf G', spełniający następujące warunki:

- $|G'| \leqslant C|G|$;
- Jeżeli $UNSAT(G) = 0$ to $UNSAT(G') = 0$;
- $UNSAT(G') \geqslant \min\{2 \cdot UNSAT(G), \alpha\}$.

Algorytm ten dla ustalonego Σ działa w czasie wielomianowym od $|G|$.

Koncepcja tu jest taka, że skoro w każdym podejściu rozmiar grafu rośnie liniowo, a $UNSAT$ się (do pewnego momentu) podwaja, to po aplikacji całej procedury $\log |G|$ razy dostaniemy $UNSAT(G') \geqslant \alpha$ lub $UNSAT(G') = 0$, bez stanów pośrednich, zaś rozmiar grafu wzrośnie wielomianowo. Gdybyśmy zatem umieli rozstrzygnąć, który z tych dwóch przypadków zachodzi, to umielibyśmy rozwiązać problem ograniczonego przypisania, o którym wiemy, że jest NP–zupełny. Przeprowadźmy formalny dowód:

Dowód PCP aproksymacyjnego z twierdzenia o amplifikacji. Dowodzimy NP–trudności problemu rozróżniania, redukujemy do problemu spełniania ograniczeń. Weźmy zatem instancję problemu spełniania ograniczeń dla, powiedzmy, $q = 2$ i $\Sigma = 3$. Jest to pewien graf G. Konstruujemy przy pomocy twierdzenia o amplifikacji ciąg grafów ograniczeń G_i, gdzie $G_0 = G$, zaś G_{i+1} to graf otrzymany z G_i przez twierdzenie o amplifikacji. Wtedy $|G_k| \leqslant C^k|G|$, oraz jeśli $UNSAT(G) = 0$ to $UNSAT(G_k) = 0$. Dodatkowo dla $t = \lceil \log \alpha \rceil$ jeśli $UNSAT(G) > 0$, to $UNSAT(G_t) \geqslant \alpha$.

Teraz jeszcze chcemy dojść z α do $1/2$. Radzimy sobie z tym prosto, kosztem podwyższenia q — niech $(1-\alpha)^u < 1/2$. Rozważmy końcowy układ ograniczeń G'', w którym $q = 2u$, zaś każdemu ciągowi u ograniczeń z G_t odpowiada jedno ograniczenie, będące AND–em tego ciągu. Wtedy jeśli w pierwotnym grafie pewne wartościowanie spełniało $1-\alpha$ spośród ograniczeń, to teraz spełnia $(1-\alpha)^u < 1/2$. Zatem zredukowaliśmy instancję spełniania ograniczeń dla $q = 2$ i $|\Sigma| = 3$ (o którym to problemie wiemy, że jest NP–trudny) do instancji rozróżniania dla $q = 2u$ i $|\Sigma| = |\Sigma_0|$. □

Może nas tu nieco martwić poszerzenie się zarówno q, jak i alfabetu. Poszerzenie się q jest w pewnym sensie nieuniknione, to cena, którą się płaci za dociągnięcie obszaru nieaproksymowalności do $1/2$ — zwróćmy uwagę, że gdyby satysfakcjonował nas stałych rozmiarów obszar nieaproksymowalności, to możemy pozostać przy $q = 2$. Przyrost rozmiaru alfabetu jest w pewnym sensie mniej bolesny — istnieją standardowe techniki teoriozłożonościowe, których nie będziemy dowodzić, które pozwalają zredukować rozmiar alfabetu do stałego rozmiaru. Mam wrażenie, że ten rozmiar, do którego możemy się zredukować, to nie więcej niż 8, ale nie chciałbym za to ręczyć. Sformułujmy tu twierdzenie, na które będziemy się powoływać:

Twierdzenie 4.8 (o kompozycji). Istnieje taka stała $\beta_3 > 0$ i taki ustalony alfabet Σ_0, że dla dowolnego grafu ograniczeń $G = (V, E, \Sigma, C)$ można w czasie liniowym od $|G|$ (dla ustalonego $|\Sigma|$) skonstruować graf $G' = (V', E', \Sigma_0, C')$, dla którego $|G'| \leqslant c(\Sigma)|G|$ oraz $\beta_3 \cdot UNSAT(G) \leqslant UNSAT(G') \leqslant UNSAT(G)$.

To twierdzenie pozostawiam bez dowodu, jest to — podobnież — stosunkowo standardowe narzędzie w teorii złożoności, mało grafowe, więc dla nas mało interesujące. Osoby ewentualnie zainteresowane odsyłam np. do pracy "The PCP theorem by Gap Amplification" Irita Dinura, sekcji 5, 7, oraz dodatku B.

4.3 Twierdzenie o amplifikacji — zarys, potrzebne składniki

Przypomnijmy — dowodzimy następującego twierdzenia o amplifikacji:

Twierdzenie 4.9. Istnieje taki alfabet Σ_0, że dla każdego Σ istnieją stałe $C > 0$ i $1 > \alpha > 0$, oraz algorytm. Algorytm bierze na wejściu graf ograniczeń $G = (V, E, \Sigma, C)$, i zwraca graf G' z ograniczeniami nad Σ_0, spełniający następujące warunki:

- $|G'| \leqslant C|G|$;
- Jeżeli $UNSAT(G) = 0$ to $UNSAT(G') = 0$;
- $UNSAT(G') \geqslant \min\{2 \cdot UNSAT(G), \alpha\}$.

Algorytm ten dla ustalonego Σ działa w czasie wielomianowym od $|G|$.

Proces amplifikacji będzie składał się z trzech części. Wpierw zrobimy preprocesing grafu, aby uczynić z niego graf d–regularny o określonej przerwie spektralnej, nie tracąc jednocześnie za dużo z warotści UNSAT. Potem będziemy nasz graf potęgować (poprawiając wartość UNSAT, acz pogarszając stopień i rozmiar alfabetu), a potem zredukujemy rozmiar alfabetu z powrotem przy pomocy twierdzenia o kompozycji. Przyjrzyjmy się przez chwilę, jak działa potęgowanie grafu. Rozważmy graf ograniczeń $G = (V, E, \Sigma, C)$, gdzie zakładamy, że G (jako graf) jest d–regularny. Definiujemy następującą operację:

Definicja 4.10. Dla $t \geqslant 1$ definiujemy potęgę grafu d–regularnego G (oznaczaną G^t) jako następujący graf ograniczeń nad alfabetem $\Sigma^{d^{\lceil t/2 \rceil}}$:

- Wierzchołki się nie zmieniają: $V(G^t) = V(G)$;
- Krawędziami G^t są ścieżki długości t w G (czyli na poziomie samego grafu jest to normalne potęgowanie);
- Alfabet ma moc $\Sigma^{d^{\lceil t/2 \rceil}}$. Pewną wartość a tego alfabetu w wierzchołku v interpretujemy jako informację o wartościach we wszystkich wierzchołkach, do których da się dojść z v w dokładnie $\lceil t/2 \rceil$ krokach. Oznaczmy zbiór tych wierzchołków przez $\Gamma(v)$. Oczywiście $|\Gamma(v)| \leqslant d^{\lceil t/2 \rceil}$, a zatem przyjmując pewien (dowolny) kanoniczny porządek możemy myśleć o wartości a jako o przypisaniu z $\Gamma(v) \to \sigma$ (ignorując nadwyżkowe wartości, jeśli jakiś wierzchołek daje się odwiedzić na więcej niż jeden sposób).
- Trzeba jeszcze zdefiniować ograniczenie na krawędzi $(u, v) \in E(G^t)$, czyli określić, które pary $(a, b) \in \Sigma^{d^{\lceil t/2 \rceil}}$ spełniają to ograniczenie. Otóż przyjmujemy jako pary spełniające takie pary a, b, że istnieje wartościowanie na $\Gamma(u) \cup \Gamma(v)$, które jest zgodne z a i b (czyli w szczególności a i b muszą być zgodne na $\Gamma(u) \cap \Gamma(v)$), oraz jest zgodne na krawędziach z $E \cap (\Gamma(u) \cup \Gamma(v))$. Czyli, innymi słowy, wartościujemy to, co a i b mówią nam, jak owartościować i sprawdzamy, czy jest to legalny fragment wartościowania na G.

Zauważmy, że potęgowanie nie psuje tego, że istnieje wartościowanie:

Fakt 4.11. Jeśli $UNSAT(G) = 0$, to $UNSAT(G^t) = 0$ dla dowolnego $t \geqslant 1$.

Bardziej ciekawy jest fakt, że tak zdefiniowana operacja, jeśli bierze graf o przyzwoitych własnościach ekspansji, to podnosi UNSAT, jeśli był on niezerowy. Kluczowym technicznym faktem będzie tu następujący lemat o amplifikacji:

Lemat 4.12 (o amplifikacji). Istnieje stała $\beta_2 = \beta(d, \Delta', |\Sigma|)$, że dla każdego d–regularnego grafu ograniczeń G, o bezwzględnej przerwie spektralnej Δ', z pętlą w każdym wierzchołku, nad alfabetem Σ zachodzi

$$UNSAT(G^t) \geqslant \beta_2 \sqrt{t} \min\{UNSAT(G), \frac{1}{t}\}.$$

Do aplikacji tego lematu potrzebny będzie pewien preprocesing:

Lemat 4.13 (o preprocesingu). Istnieją takie stałe C, $d > 2$, $\beta_1 > 0$ i $\Delta' > 0$, że dla dowolnego grafu ograniczeń G w czasie wielomianowym od $|G|$ da się skonstruować graf ograniczeń G' spełniający:

- G' jest d–regularny i ma pętlę w każdym wierzchołku;
- $|G'| \leqslant C|G|$;

- $\Delta'(G') \geqslant \Delta'$;
- G' ma ten sam alfabet co G;
- $\beta_1 \cdot UNSAT(G) \leqslant UNSAT(G') \leqslant UNSAT(G)$.

Zakładając te dwa lematy będziemy w stanie udowodnić twierdzenie o amplifikacji (a więc i całe PCP):

Dowód. Weźmy graf G. Niech β_1 będzie dane przez lemat o preprocesingu, β_3 — przez twierdzenie o kompozycji, niech $\beta_2 = \beta(d, \Delta', |\Sigma|)$, gdzie d i Δ' pochodzą z lematu o preprocesingu, zaś Σ to alfabet grafu G. Weźmy $t \geqslant (2/\beta_1\beta_2\beta_3)^2$.

Niech G_1 będzie grafem z lematu o preprocesingu dla G, $G_2 = G_1^t$, zaś G_3 — grafem z twierdzenia o kompozycji dla G_2. Sprawdzamy po kolei:

- Rozmiar grafu w lemacie o preprocesingu oraz w twierdzeniu o kompozycji rośnie liniowo. Przy potęgowaniu liczba wierzchołków nie rośnie, stopień każdego wierzchołka to d^t, zatem liczba krawędzi jest liniowa od liczby wierzchołków, skąd w sumie $|G_3| \leqslant C|G|$;
- Wszystkie grafy są liniowego rozmiaru, a algorytmy — wielomianowe od rozmiaru grafu, czyli cały proces działa wielomianowo;
- Jeśli $UNSAT(G) = 0$, to żadna z trzech operacji tego nie zmienia;
- Jeśli natomiast $UNSAT(G) > 0$, to $UNSAT(G_3) \geqslant \beta_3 \min\{\beta_2\beta_1\sqrt{t}UNSAT(G), 1/t\}$, czyli dla $\alpha = \beta_3/t$ teza twierdzenia o amplifikacji zachodzi.

□

4.4 Twierdzenie PCP — preprocesing

Teraz udowodnimy lemat o preprocesingu. Przypomnijmy — chcemy dowolny graf zmienić w graf d–regularny z pętlami, stałą przerwą spektralną, o liniowo dużym rozmiarze, z kontrolą nad zmianą parametru $UNSAT$. Zrobimy to w dwóch ruchach.

4.4.1 Rozbicie wierzchołków

Ustalmy sobie rodzinę ekspanderów o stopniu $d-1$ i bezwzględnej przerwie spektralnej nie mniejszej niż $\Delta' > 0$, taką, że każdy ekspander konstruuje się w czasie wielomianowym od swojego rozmiaru. Na wykładzie o zygzaku pokazaliśmy, jak dostać rodzinę dla $n = (d-1)^{8t}$, dostanie rodziny działającej dla dowolnego n (być może z nieco gorszym Δ') jest nietrudnym ćwiczeniem. Z twierdzeń o porównywaniu przerw i ekspansji wynika w szczególności, że wszystkie te ekspandery mają ekspansję krawędziową nie mniejszą niż pewne $h > 0$.

Rozważmy graf ograniczeń G. Wpierw definiujemy graf G' następująco — każdy wierzchołek $v \in V(G)$ zmieniamy na $\deg v$ wierzchołków $[v]$ w grafie G'. Dla krawędzi $uv \in E(G)$ dodajemy dokładnie jedną krawędź pomiędzy pewnym wierzchołkiem z $[v]$ i pewnym wierzchołkiem z $[u]$, każdy wierzchołek ma dokładnie jedną taką krawędź zewnątrzną (obrazek jest taki, jak u Marcina w rysunku zygzaku). Na tej krawędzi kładziemy ograniczenie takie, jak na pierwotnym uv. Dodatkowo na $[v]$ kładziemy krawędzie wedle ekspandera z naszej rodziny o $\deg v$ wierzchołkach, zaś jako ograniczenia kładziemy równości.

Graf, który otrzymaliśmy jest oczywiście d–regularny. Nie ma jeszcze pętli ani sensownej przerwy spektralnej. Jego rozmiar wzrósł, wbrew pozorom, liniowo — każda stara krawędź zrodziła do dwóch nowych wierzchołków (jeden, jeśli była pętlą), zaś nowych krawędzi jest liniowo wiele (bo graf jest regularny). Sprawdźmy, co się stało z parametrem $UNSAT$.

Stwierdzenie 4.14. $cUNSAT(G) \leqslant UNSAT(G') \leqslant UNSAT(G)$.

Dowód. Wpierw prosta część — $UNSAT$ nie wzrósł. Weźmy optymalne przypisanie ϕ dla G. W sposób naturalny daje ono przypisanie ϕ' na G' — na wierzchołkach z $[v]$ kładziemy $\phi(v)$. Wtedy ograniczenia na krawędziach "wewnętrznych" są spełnione, zaś na zewnętrznych są spełniowe wtedy i tylko wtedy, jeśli były spełnione w G. Czyli na sztuki niespełnionych ograniczeń jest tyle samo, a krawędzi przybyło — zatem $UNSAT$ nie wzrósł (a nawet, zapewne, zmalał).

Teraz trudniejsza część. Idea jest taka, że "gęsta" struktura ekspandera w chmurze wierzchołków będzie silnie karała za przypisywanie wierzchołkom wewnętrz chmury różnych wartości.

To teraz formalnie. Weźmy dowolne przypisanie σ' w G'. Weźmy przypisanie σ w G, które jest przypisaniem większościowym — tj. dla każdego wierzchołka $v \in V(G)$ wybieramy ten element Σ, który jest najliczniej reprezentowany w $[v]$ przez σ'. Zauważmy, że $|E(G')| \leqslant d|E|$. Niech F będzie zbiorem tych krawędzi w E, których ograniczenia nie są spełnione przez σ, wtedy $|F| \geqslant \alpha|E| \geqslant \alpha|E'|/d$. Analogicznie definiujemy F' jako zbiór tych krawędzi z E (nie z E'), których ograniczenia nie są spełnione przez σ'. Niech S będzie zbiorem tych wierzchołków z V', które nie są zgodne z najliczniejszym przypisaniem w swojej grupie wierzchołków.

Zauważmy, że każda krawędź z F ma albo przynajmniej jeden koniec w S, albo należy do F'. Stąd $|F'|+|S| \geqslant \alpha|E'|/d$. Jeśli $|F'| \geqslant \alpha|E'|/2d$, to już koniec — mamy $UNSAT(\sigma')$ ograniczony z dołu przez pewną stałą. W przeciwnym razie dzielimy S na zbiory $S(v,a)$, gdzie $v \in V(G)$, $a \in \Sigma$ — $S(v,a)$ to zbiór tych wierzchołków z $[v]$, dla których σ' przyjmuje wartość a. Ze zbioru $S(v,a)$ wychodzi w $[v]$ przynajmniej $h|S(v,a)|$ krawędzi — jako, że a nie było najliczniejszym przypisaniem w $[v]$, to $|S(v,a)| \leqslant [v]/2$. Każda z tych krawędzi miała ograniczenie równościowe, które — jako, że krawędź opuszcza $S(v,a)$ — nie jest spełnione. Zatem, sumując po wszystkich v i a, otrzymujemy przynajmniej $h|S|/2$ krawędzi nie spełniających swoich ograniczeń, co — wobec założenia $|S| \geqslant |E'|/2d$ — znowu kończy zadanie. □

4.4.2 Dodanie ekspandera

Udało nam się zatem (bardzo korzystając z ekspanderów) zbić stopień nie tracąc jednocześnie zbyt wiele na $UNSAT$. Teraz chcemy jeszcze podbić ekspansję i dodać pętle. To zrobimy bardzo prosto — pętle po prostu dodamy, oraz, dodatkowo, weźmiemy ekspander z naszej rodziny na $V(G)$ i dodamy jego krawędzie (z krotnościami) z pustymi (tj. zawsze spełnionymi) ograniczeniami.

Rozmiar naszego grafu wzrósł najwyżej dwukrotnie (bo dwukrotnie zwiększyła się liczba krawędzi). Zauważmy też, że graf docelowy ma pętle i jest $2d$–regularny. Co więcej, ekspansja wierzchołkowa nowego grafu jest nie gorsza niż ekspansja wierzchołkowa naszego ekspandera z rodziny (bo dodanie krawędzi nie pogarsza ekspansji wierzchołkowej), a zatem, skoro mamy kontrolę nad stopniem i ekspansją wierzchołkową, to kontrolujemy też przerwę spektralną wynikowego grafu. I na koniec $UNSAT$ nie podniósł się (bo nowe krawędzie zawsze są spełnione, więc każde przypisanie ma lepszy mniejszy $UNSAT$ w nowym grafie niż w starym), a spadł co najwyżej dwukrotnie (bo dodaliśmy drugie tyle krawędzi).

To kończy dowód lematu o preprocesingu — w wyniku dwóch przeprowadzonych transformacji graf wynikowy ma stopień $2d$, ograniczoną od dołu przerwę spektralną, $UNSAT$ nie wzrósł, a spadł co najwyżej o stałą frakcję, rozmiar wzrósł liniowo, alfabet się nie zmienił, zaś cała konstrukcja była wielomianowa (suma wielomianów od stopni w pierwszym ruchu, wielomian od rozmiaru w drugim na konstrukcję ekspanderów, oraz trywialnie wielomianowe pozostałe operacje, np. sumowanie grafów i dodawanie pętli).

4.5 Lemat o amplifikacji

Dowodzimy następującego lematu:

Lemat 4.15 (o amplifikacji). Istnieje stała $\beta_2 = \beta(d, \Delta', |\Sigma|)$, że dla każdego d–regularnego grafu ograniczeń G, o bezwzględnej przerwie spektralnej Δ', z pętlą w każdym wierzchołku, nad alfabetem Σ zachodzi

$$UNSAT(G^t) \geqslant \beta_2\sqrt{t}\min\{UNSAT(G), \frac{1}{t}\}.$$

Załóżmy dla prostoty zapisu, że t jest liczbą parzystą (w ten sposób nie będziemy wszędzie pisać sufitów). Oznaczmy $d^{t/2}$ przez D.

Udowodnimy nawet silniejszy nieco fakt. Weźmy dowolne wartościowanie $\bar{\sigma} : V(G)\to\Sigma^D$. Ustalmy wartościowanie większościowe $\sigma : V(G)\to\Sigma$, które wierzchołkowi v przypisuje wartość większościową z $\bar{\sigma}$ — puszczamy błądzenie losowe długości $t/2$ z v, lądujemy w w, patrzymy, jaki kolor wartościowanie $\bar{\sigma}(w)$ przypisuje wierzchołkowi v (oznaczać to będziemy przez $(\bar{\sigma}(w))_v$) i wybieramy jako $\sigma(v)$ ten kolor, który ma w tym procesie najwyższe prawdopodobieństwo. Udowodnimy, że $UNSAT(\bar{\sigma}) \geqslant \beta_2\sqrt{t}\min\{UNSAT(\sigma), 1/t\}$, biorąc minimum po $\bar{\sigma}$ dostaniemy tezę.

Wpierw intuicja, czemu to działa. Weźmy wartościowanie σ, i załóżmy, że wartościowanie $\bar{\sigma}$ pochodzi od tego σ. Jeśli wybierzemy jedną, losową, krawędź, to prawdopodobieństwo, że jej ograniczenie nie jest spełnione to $UNSAT(\sigma)$. Jeśli weźmiemy t losowych krawędzi, to to prawdopodobieństwo powinno wzrosnąć około t razy (jeśli $UNSAT(\sigma)$ jest małe w porównaniu do $1/t$). Idea jest taka, że branie losowego błądzenia po ekspanderze powinno dobrze symulować losowy wybór krawędzi. Zatem łatwo uwierzyć, że dla wartościowań pochodzących od pewnego wartościowania pierwotnego faktycznie powinniśmy mieć dobre oszacowanie. Trik teraz polega na poradzeniu sobie z wartościowaniami, które nie pochodzą od niczego. Nasza nadzieja jest taka, że one przynajmniej w racjonalnym stopniu "pochodzą" od swojego większościowego wartościowania.

Zacznijmy od ustalenia zbioru krawędzi F, dla których σ nie spełnia ograniczeń. Jeśli $UNSAT(\sigma) < 1/t$, to przyjmujemy za F wszystkie takie krawędzie, jeśli nie, to przyjmujemy dowolny zbiór $|E|/t$ takich krawędzi. Teraz $|F|/|E| \leqslant \min\{UNSAT(\sigma), 1/t\}$, i z dokładnością do błędu zaokrągleń zachodzi równość.

Przypomnijmy, że krawędzie w G^t odpowiadały ścieżkom długości t w G. Nieco nadużywając notacji będziemy mówić, że ścieżka $(v_0, v_1, \ldots, v_t)$ jest krawędzią w G^t. Powiemy, że ścieżka $(v_0, v_1, \ldots, v_t)$ jest zbita na i–tej krawędzi, jeśli $(v_{i-1}, v_i) \in F$ i $\bar{\sigma}(v_0)_{v_{i-1}} = \sigma(v_{i-1})$ oraz $\bar{\sigma}(v_t)_{v_i} = \sigma(v_i)$ (czyli opinia v_0 o v_{i-1} oraz v_t o v_i zgadzają się z większościową). Tu uwaga — w szczególności zakładamy, że te opinie istnieją. Oczywiście jeśli ścieżka jest bita przez jakąkolwiek swoją krawędź, to $\bar{\sigma}$ nie spełnia ograniczeń G^t na tej ścieżce.

Niech $\mathbf{e}$ będzie losową ścieżką. Zwróćmy uwagę, że stopień każdego wierzchołka jest taki sam, więc losowa ścieżka zaczyna się od losowego wierzchołka. Przez $N(\mathbf{e})$ oznaczmy liczbę krawędzi, które biją $\mathbf{e}$ spośród $t/2-\sqrt{t}, t/2+\sqrt{t}$ (czyli interesują nas tylko środkowe krawędzie). Interesować nas będzie $\mathbb{P}(N(\mathbf{e}) > 0)$ (widać, że szacujemy bardzo na pałę, i stąd weźmie się w ostatecznym szacowaniu ten $\sqrt{t}$ — da się wyciągnąć i t, ale nam to nie jest potrzebne).

Będziemy liczyć wartości oczekiwane (bo to się prościej liczy). Zauważmy, że jeśli $X \geqslant 0$, to

$$\mathbb{E}X = \mathbb{E}X\mathbf{1}_{X>0} \leqslant \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{P}(X > 0)},$$

skąd $\mathbb{P}(N(\mathbf{e}) > 0) \geqslant \frac{(\mathbb{E}N(\mathbf{e}))^2}{\mathbb{E}N^2(\mathbf{e})}$.

4.6 Liczymy $\mathbb{E}N$

Wpierw liczymy $\mathbb{E}N$. Ustalmy numer krawędzi — $\mathbb{E}N$ to suma prawodopodobieństw, że i–ta krawędź bije ścieżkę. Ustalmy dowolną krawędź z F — ta krawędź jest i–tą krawędzią ścieżki z prawdopodobieństwem $1/|E|$, bo graf jest d–regularny, więc z każdej krawędzi grafu jesteśmy w stanie ustalić tyle samo ścieżek, mających daną krawędź jaką i–tą. Czyli interesuje nas jeszcze tylko to, żeby opinia v_0 o v_{i-1} oraz v_t o v_i zgadzały się z σ (zwróćmy uwagę, że a priori te opinie nie muszą istnieć, zauważmy też, że ignorujemy tu opinie v_0 o v_i oraz v_t o v_{i-1}, które mogłyby nam pomóc).

Czyli tak — puszczamy z v_{i-1} o $i-1$ krawędziach i pytamy, jakie jest prawdopodobieństwo, że dojdziemy do wierzchołka w, w którym $\bar{\sigma}(w)_{v_{i-1}} = \sigma(v_{i-1})$. Zauważmy, że jeśli $i-1 = t/2$, to to prawdopodobieństwo jest równe przynajmniej $1/|\Sigma|$ — bo $\sigma(v_{i-1})$ zostało wybrane jako najliczniejsze spośród przypisań na wierzchołkach odległych o dokładnie $t/2$ od v_{i-1}.

Teraz jest fragment probabilistyczny, który pominę, choć nie jest trudny. Pomysł jest taki, że skoro w każdym wierzchołku jest pętla, to tak naprawdę ścieżki długości l to tak naprawdę ścieżki długości około $l(d-1)/d$, oraz pętle. I, co więcej, jak weźmiemy ścieżkę długości l bliskiej $t/2$, to prawdopodobieństwo tego, że wykonamy na niej dokładnie k ruchów w przód (nie po pętli) jest bliskie prawdopodobieństwu tego, że wykonamy k ruchów na ścieżce $t/2$, o ile k jest bliskie wartości oczekiwanej:

Lemat 4.16. Jeśli mamy dwa rozkładny dwumienne z tym samym prawdopodobieństwem sukcesu p, i $l_0 - \sqrt{l_0} \leqslant l_1 \leqslant l_0 + \sqrt{l_0}$ oraz stałą c, to istnieje taka stała $\tau(c,d)$, że dla dostatecznie dużych l_0 i każdego k spełniającego $|k - pl_0| < c\sqrt{l_0}$ zachodzi

$$\tau \leqslant \frac{\mathbb{P}(B_{l_0,p} = k)}{\mathbb{P}(B_{l_1,p} = k)} \leqslant \frac{1}{\tau}.$$

To, co mówi ten fakcik na nasze, to że jeśli długości ścieżek są bliskie (w promieniu pierwiastka jednej z nich), to prawdopodobieństwa tego, że obie te ścieżki będą miały pewną ustaloną liczbę niepętlowych krawędzi jest podobne, o ile ta liczba jest bliska wartości oczekiwanej (znowu, w promieniu pierwiastkowym).

To teraz dlaczego stąd wynika teza (gdzie tezą jest, że tam jest sporo takich ścieżek)? Otóż chcemy policzyć, ile jest ścieżek długości $i-1$, które kończą bieg w takim wierzchołku, który przypisuje do v_{i-1} wartość $\sigma(v_{i-1})$ (czyli tę większościową). Otóż jest ich

$$\sum_{k=0}^{i-1} \mathbb{P}(k \text{ kroków z } i-1)\mathbb{P}(\text{ścieżka dł. } k \text{ trafia dobrze}).$$

Tę sumę szacujemy przez sumę po okolicach wartości oczekiwanej k, czyli $k_0 = (i-1)(d-1)/d$:

$$\sum_{k=k_0-C\sqrt{k_0}}^{k_0+C\sqrt{k_0}} \mathbb{P}(k \text{ kroków z } i-1)\mathbb{P}(\text{ścieżka dł. } k \text{ trafia dobrze}).$$

Na tych wartościach k prawdopodobieństwo trafienia dobrze k kroków z $i-1$ jest bliskie prawdopodobieństwu trafienia dobrze k kroków z $t/2$, szacujemy z dołu przez:

$$\sum_{k=k_0-C\sqrt{k_0}}^{k_0+C\sqrt{k_0}} \tau^{-1}\mathbb{P}(k \text{ kroków z } t/2)\mathbb{P}(\text{ścieżka dł. } k \text{ trafia dobrze}).$$

Stała τ zależy od C, nie przeszkadza nam to. Teraz dodajemy z powrotem resztę sumy, i odejmujemy:

$$\tau^{-1}\sum_{k=1}^{t/} \mathbb{P}(k \text{ kroków z } t/2)\mathbb{P}(\text{ścieżka dł. } k \text{ trafia dobrze}) - \tau^{-1} \sum_{k \notin [k_0 \pm \sqrt{k_0}]} \mathbb{P}(k \text{ kroków z } t/2).$$

Pierwotna suma to przynajmniej $1/|\Sigma|$, czyli razem po lewej dostajemy $(\tau(C)|\Sigma|)^{-1}$. Teraz dobieramy C tak, żeby prawa strona była mniejsza od $(2|\Sigma|)^{-1}$, i już jest dobrze — dostajemy szacowanie całości z dołu przez $(2\tau|\Sigma|)^{-1}$, czyli tyle, ile oczekujemy. Sumując po i dostaniemy $\mathbb{E}N \geqslant C\sqrt{t}|F|/|E|$.

4.7 Liczymy $\mathbb{E}N^2$

To będzie istotnie prostsze — oszacujemy od góry ile razy losowa ścieżka przecina zbiór F — to oczywiście szacuje z góry $N(\mathbf{e})$ (czyli w ogólnie nie martwimy się przypisaniami). Jak poprzednio piszemy $Z(\mathbf{e}) = \sum Z_i(\mathbf{e})$ — odpowiednie indykatory trafienia na i–tej krawędzi. Oczywiście $\mathbb{E}N^2(\mathbf{e}) \leqslant \mathbb{E}Z^2(\mathbf{e}) = \sum \mathbb{E}Z_i^2(\mathbf{e}) + 2\sum \mathbb{E}Z_iZ_j = |I|\frac{|F|}{|E|} + 2\sum \mathbb{E}Z_iZ_j$, gdzie I to odcinek $t/2 \pm \sqrt{t}$.

Teraz koncept jest taki, że ponieważ graf jest ekspanderem, to korelacje pomiędzy krawędziami nie trwają zbyt długo. Przypomnijmy, że na ćwiczeniach był taki fakt (zadanie 74), że jeśli zaczynam w złej krawędzi i wykonuję i kroków, to i–ty wykonam złą krawędzią z prawdopodobieństwem co najwyżej $\frac{|F|}{|E|} + ((d-\Delta')/d)^i$. Czyli tu $\mathbb{E}Z_iZ_j = \mathbb{P}(Z_i = 1)\mathbb{P}(Z_j = 1|Z_i = 1) \leqslant \frac{|F|}{|E|}(\frac{|F|}{|E|} + \lambda^{|j-i|-2})$, gdzie $\lambda = (d-\Delta')/d$ jest jakąś stałą. Wobec tego mamy $\sum_{i<j} \mathbb{E}Z_iZ_j = |I|(|I|-1)\frac{|F|^2}{|E|^2} + \frac{|F|}{|E|}\sum_{i\in I}\sum_{j>i}\lambda^{j-i-2}$. Wewnętrzna suma zwija się do stałej, czyli dostajemy $|I|\frac{|F|}{|E|}$. Teraz $|I|\frac{|F|}{|E|} < 1$, skąd całość szacuje się z góry przez $C\sqrt{t}\frac{|F|}{|E|}$.

Wobec tego, szacując całość, dostajemy $\mathbb{P}(N(\mathbf{e}) > 0) \geqslant \frac{(\mathbb{E}N)^2}{\mathbb{E}N^2} \geqslant C\sqrt{t}\frac{|F|}{|E|}$, co kończy dowód lematu o amplifikacji.