Informatyczny kącik olimpijski (9) – zliczamy ścieżki

W kolejnym kąciku olimpijskim zmierzymy się z takim zadaniem:

Mamy dany nieskierowany graf prosty G (bez pętli i wielokrotnych krawędzi). Zadaniem jest obliczenie liczby ścieżek prostych długości k w tym grafie.

Zadanie to ma wiele podobnych wersji – liczba k może być ustalona, zamiast ścieżek prostych możemy liczyć cykle proste (dla przypomnienia: ścieżka prosta nie przechodzi przez żaden wierzchołek więcej niż raz, cykl prosty podobnie), itd.

Najprostsze rozwiązanie jest dość oczywiste – będziemy *explicite* poszukiwać takich ścieżek. Z każdego wierzchołka będziemy uruchamiać algorytm DFS, z zadaniem szukania ścieżek długości k, z każdego następnika będziemy szukać ścieżek długości $k-1$ itd. Wchodząc do wierzchołka będziemy go zaznaczać jako odwiedzony (żeby nie powtarzać wierzchołków na ścieżkach), a wychodząc – z powrotem jako nieodwiedzony (żeby rozróżnić np. ścieżki $v, x, w, \ldots$ i $v, y, w, \ldots$). Taki algorytm jest bardzo prosty do zapisania, zużywa niewiele pamięci, niestety jego złożoność czasowa jest rzędu n^{k+1}.

Żeby go usprawnić, rozważmy poszczególne wartości k.

Dla $k = 2$ pozostańmy przy rozwiązaniu o złożoności $O(n^3)$. Na potrzeby późniejszych rozważań zamiast korzystać z algorytmu DFS, obliczymy tablicę T_2, taką, że $T_2[a, b]$ jest liczbą ścieżek prostych prowadzących z a do b o długości 2 o ile $a \neq b$, a jeśli $a = b$, to $T_2[a, b]$ jest liczbą cykli prostych o długości 2 przechodzących przez a (czyli, *de facto*, liczbą krawędzi wychodzących z a). Wtedy rozwiązaniem jest suma $\sum_{a<b} T_2[a, b]$, zaś tablicę T_2 obliczamy oczywiście w sposób następujący:

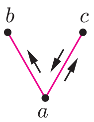
$$T_2[a, b] = \sum_{c} T_1[a, c] \cdot T_1[c, b],$$

gdzie $T_1[a, b]$ jest równe 1, jeśli z a do b istnieje krawędź, zaś 0 w przeciwnym przypadku.

Ten sam sposób rozumowania przyspiesza już nam o rząd wielkości rozwiązanie problemu dla $k = 3$. Znając T_1, obliczamy T_2 w czasie $O(n^3)$, natomiast

$$T_3[a, b] = \sum_{c \neq b} T_1[a, c] \cdot (T_2[c, b] - T_1[a, b]).$$

Dlaczego odejmujemy $T_1[a, b]$? Wszystko wyjaśnia poniższy rysunek:

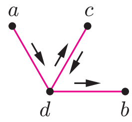


Wraz ze zwiększającą się długością ścieżki, i ułatwieniami które stosujemy, pojawiają się ścieżki które nie są proste, a które mają taką strukturę, że moglibyśmy je przypadkiem policzyć, gdybyśmy napisali po prostu $T_1[a, c] \cdot T_2[c, b]$. Podobnie będzie w następnym przypadku.

Kiedy $k = 4$, narzuca się, żebyśmy skorzystali dwa razy z tablicy T_2, a więc napisali coś takiego:

$$T_4[a, b] = \sum_{c \neq a, b} T_2[a, c] \cdot T_2[c, b] \qquad \text{(na pewno?)}.$$

Pojawia się jednak podobny problem jak w poprzednim przypadku, przedstawiony na rysunku:



Możemy sobie z nim poradzić na dwa sposoby:

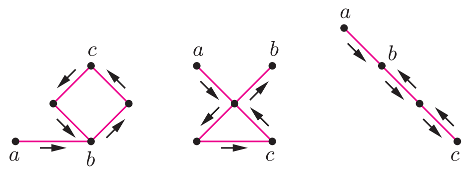
• Od wartości $T_4[a, b]$ odejmijmy jeszcze sumę $\sum_{d \neq a, b} T_1[a, d] \cdot T_1[d, b] \cdot (deg(d) - 2)$. Zmienna d odpowiada wierzchołkowi d z rysunku. W ten sposób odejmujemy każdy taki przypadek.

• Pozostawmy tablicę T_4 taką, jaka jest: dopuszcza ona pewne ścieżki nie-proste, ale nas interesuje liczba ścieżek prostych długości 4 w całym grafie. Wtedy odpowiedzią będzie, jak w poprzednich przypadkach, suma $\sum_{a<b} T_4[a, b]$, pomniejszona o liczbę „rozetek” w grafie (takich właśnie, jak na rysunku), równą $\sum_d \frac{deg(d) \cdot (deg(d)-1) \cdot (deg(d)-2)}{2}$. Dzielimy przez 2, a nie przez 6, ponieważ każdą „rozetkę” odejmujemy 3 razy – przy liczeniu ścieżek z a do b, z a do c i z c do b.

Tym samym dochodzimy do ostatniego przypadku, którym się zajmiemy – $k = 5$. Tutaj rozważania stają się dość skomplikowane. Poprawnym rozwiązaniem jest zastosowanie podobnych rozważań jak wcześniej, zaczynając od napisania

$$T_5[a, b] = \sum_{c} T_3[a, c] \cdot T_2[c, b] \qquad \text{(na pewno?)}.$$

Tutaj jednak pojawia się wiele przypadków specjalnych, które należy odjąć. Oto niektóre z nich:



Najłatwiej tym razem postąpić podobnie, jak w przypadku drugim w rozwiązaniu dla $k = 4$ – tablicę T_5 pozostawić taką, jaka jest, a przypadki szczególne zliczać globalnie dla całego grafu.

Czytelnikowi, który chciałby się zmierzyć z dokładnym rozwiązaniem, i implementacją powyższych zadań dla $k = 4$ i $k = 5$, polecam zadania E – „Piąty wymiar” i G – „Autostrady” z Akademickich Mistrzostw Polski w Programowaniu Zespołowym 2007.

Polecam również zastanowienie się nad podobnym podejściem do innych zadań tej samej klasy – wymienionych na wstępie lub np. do następującego problemu: *Mamy dany zbiór punktów na płaszczyźnie, z których żadne 3 nie są współliniowe. Ile tworzą one wypukłych czworokątów?*

Filip WOLSKI