

Informatyczny kącik olimpijski (105): Odtwarzanie drzewa

W tym miesiącu omówimy zadanie *New Year and Forgotten Tree*, które pojawiło się na konkursie Good Bye 2015 na platformie Codeforces.

Standardowym opisem nieskierowanego drzewa o n wierzchołkach (ponumerowanych od 1 do n) jest podanie $n - 1$ par liczb reprezentujących jego krawędzie. Na wejściu analizowanego zadania dany jest taki właśnie opis, jednakże każda cyfra została zastąpiona znakiem '?'. Na przykład, krawędź między wierzchołkami 90 i 123 byłaby reprezentowana jako „?? ???”. Naszym zadaniem jest odtworzyć jakiegokolwiek drzewo pasujące do danego opisu bądź stwierdzić, że takie drzewo nie istnieje.

Niech k oznacza największą możliwą liczbę cyfr etykiety wierzchołka (czyli $k \approx \log_{10} n$). Wierzchołki o tej samej liczbie cyfr są, oczywiście, reprezentowane przez ten sam napis, więc w naturalny sposób podzielimy wierzchołki na k grup, zgodnie z liczbą cyfr etykiety wierzchołka. Dla każdej pary takich grup znamy żadaną liczbę krawędzi między wierzchołkami tych grup.

Okazuje się, że próba prostego zachłannego odtwarzania naszego drzewa nie rozwiązuje zadania. Również (być może narzucające się) użycie twierdzenia Halla (dla 2^k zbiorów grup) też nie jest poprawne. Dobrym pomysłem jest natomiast konstrukcja, w której utrzymujemy cały czas jedną spójną składową i dokładamy do niej nowe krawędzie i wierzchołki.

Spróbujemy w każdej grupie wyróżnić jeden wierzchołek w sposób prowadzący do pewnych dwóch własności. Po pierwsze, chcemy, by krawędzie między wyróżnionymi wierzchołkami łączyły je w jedną spójną składową. Równoważnie, między k wyróżnionymi wierzchołkami potrzebujemy $k - 1$ krawędzi.

Rozważmy dowolne rozwiązanie (drzewo pasujące do danego opisu). Wybierzmy jedną dowolną krawędź między dwoma wierzchołkami z różnych grup i wyróżnijmy te dwa wierzchołki. Dopóki wyróżnionych jest mniej niż k wierzchołków, powtarzamy następujące kroki. Grupę nazwijmy nową, jeśli nie zawiera jeszcze żadnego wyróżnionego wierzchołka. Wierzchołek nazwijmy osiągalnym, jeśli da się do niego dojść z jednego z już wyróżnionych wierzchołków bez przechodzenia przez wierzchołki z nowych grup. Z takich definicji wynika, że istnieje krawędź między osiągalnym wierzchołkiem i wierzchołkiem z nowej grupy. Te dwa wierzchołki oznaczmy jako u i v . Jeśli u jest wyróżnione, to po prostu wyróżniamy też v . W przeciwnym przypadku zastąpmy najpierw krawędź między u i v krawędzią między v i wierzchołkiem wyróżnionym w grupie z u (nazwijmy go u'). Powstały graf wciąż pasuje do tego samego opisu oraz wciąż jest drzewem, bo po usunięciu krawędzi uv wierzchołki u i u' są w jednej składowej (bo u było osiągalne). Teraz możemy wyróżnić v i wciąż spełniony jest żądany warunek, że krawędzie między wyróżnionymi wierzchołkami łączą je w jedną składową.

Po drugie, niech każda krawędź w drzewie dotyka co najmniej jednego wyróżnionego wierzchołka. Jeśli tak nie jest, to spójrzmy na dowolną krawędź uv między dwoma niewyróżnionymi wierzchołkami z różnych grup. Jako u' i v' oznaczmy wyróżnione wierzchołki z grup zawierających odpowiednio u i v . Usuńmy krawędź uv , co dzieli drzewo na dwie składowe. Wiemy, że wyróżnione wierzchołki są ze sobą połączone w jedną składową, zatem u' i v' są wciąż w jednej składowej. Po usunięciu krawędzi uv wierzchołki u i v muszą

być w dwóch różnych składowych, zatem jeden z nich musi być w tej samej składowej co u' i v' .

Bez straty ogólności powiedzmy, że wierzchołek u jest w tej samej składowej co u' i v' . Zatem wierzchołek v jest w innej składowej i możemy go połączyć nową krawędzią z u' . Nowe drzewo wciąż pasuje do danego opisu. Podobne rozumowanie można zastosować dla krawędzi między dwoma niewyróżnionymi wierzchołkami z tej samej grupy, co pozostawiamy Czytelnikowi jako proste ćwiczenie. Skoro możemy pozbywać się krawędzi między niewyróżnionymi wierzchołkami, to musi istnieć rozwiązanie, które takich krawędzi nie ma wcale.

Pokazaliśmy, że jeśli istnieje jakiegokolwiek rozwiązanie, to istnieje też takie, w którym da się wyróżnić po jednym wierzchołku w każdej z k grup w taki sposób, że pewien podgraf drzewa jest drzewem rozpinającym k wyróżnionych wierzchołków oraz że nie ma krawędzi między niewyróżnionymi wierzchołkami. Możemy bez straty ogólności powiedzieć, że wyróżniamy wierzchołki o numerach 10^i . Skoro k jest bardzo małe, to po prostu rozważmy kolejno wszystkie k^{k-2} drzewa rozpinające o tylu wierzchołkach. W zadaniu tym można to zrobić brutalnie na różne sposoby, natomiast zainteresowanym Czytelnikom polecamy zapoznanie się z kodami Prüfera, które opisują wydajną metodę iteracji po wszystkich etykietowanych drzewach o ustalonej liczbie wierzchołków.

Powiedzmy, że ustaliliśmy $k - 1$ krawędzi między k wyróżnionymi wierzchołkami. Pozostałe $n - 1 - (k - 1) = n - k$ krawędzi nazwijmy zwykłymi. Zwykła krawędź dotyka dokładnie jednego wyróżnionego i jednego niewyróżnionego wierzchołka. Każda taka krawędź musi dotykać innego niewyróżnionego wierzchołka, bo takich wierzchołków jest $n - k$, czyli tyle, ile jest zwykłych krawędzi. Kolejność niewyróżnionych wierzchołków w grupie nie ma znaczenia, bo każdy z nich jest reprezentowany przez ten sam napis na wejściu. Pozostaje zatem dla każdej takiej krawędzi zdecydować, w której z dwóch grup ma ona dotykać wyróżnionego wierzchołka, a w której ma dotykać nowego niewyróżnionego wierzchołka. Pamiętajmy, że w każdej grupie znamy docelową liczbę niewyróżnionych wierzchołków.

Tę ostatnią część zadania możemy rozwiązać za pomocą algorytmu maksymalnego przepływu. Tworzymy k wierzchołków g_1, \dots, g_k reprezentujących grupy. Prowadzimy z nich do ujścia krawędzie o przepustowościach równych liczbom niewyróżnionych wierzchołków w odpowiednich grupach. Tworzymy też $\frac{k \cdot (k+1)}{2}$ wierzchołków $e_{i,j}$, do których ze źródła prowadzimy krawędzie o przepustowościach równych liczbom zwykłych krawędzi między grupami i i j . Z $e_{i,j}$ prowadzimy krawędzie do g_i i do g_j o nieskończonych przepustowościach. Przepływ o wartości x z $e_{i,j}$ do g_i należy interpretować jako decyzję, by x krawędzi między grupami i i j dotykało niewyróżnionych wierzchołków w i .

Rozważyć musimy wszystkie k^{k-2} drzew rozpinających na k wierzchołkach i dla każdego znaleźć przepływ w grafie o $O(k^2)$ wierzchołkach i krawędziach. Przykładowa złożoność takiego programu to $O(k^{k-2} \cdot k^6 + n) = O(k^{k+4} + n)$, gdzie obecność składnika n wynika z konieczności właściwego odtworzenia i wypisania drzewa o n wierzchołkach.

Zainteresowanym proponujemy zastanowić się, czy możemy zamiast metody maksymalnego przepływu użyć twierdzenia Halla.

Kamil DĘBOWSKI