



Politechnika Wroclawska

Algorytmy wyznaczania centralności w sieci

Szymon Szyłko

*Zakład systemów Informacyjnych
Wrocław 10.01.2008*





Agenda prezentacji

- Cechy sieci
- Algorytmy grafowe
- Badanie centralności
- Algorytmy wyznaczania centralności
- Przybliżone metody wyznaczania centralności



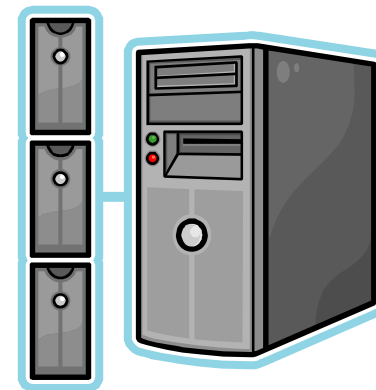
Cechy sieci

- Globalne własności sieci
 - Średnica sieci
 - Stopień klasteryzacji
 - Stopień spójności
 - Wyznaczanie maksymalnego przepływu
- Lokalne (centralność węzła, krawędzi)
 - Wyznaczanie węzłów kluczowych
 - Wyznaczanie stopnia powiązań
 - Wyznaczanie centralności jako środka ciężkości
 - Wyznaczanie miary „betweennes”



Algorytmy grafowe

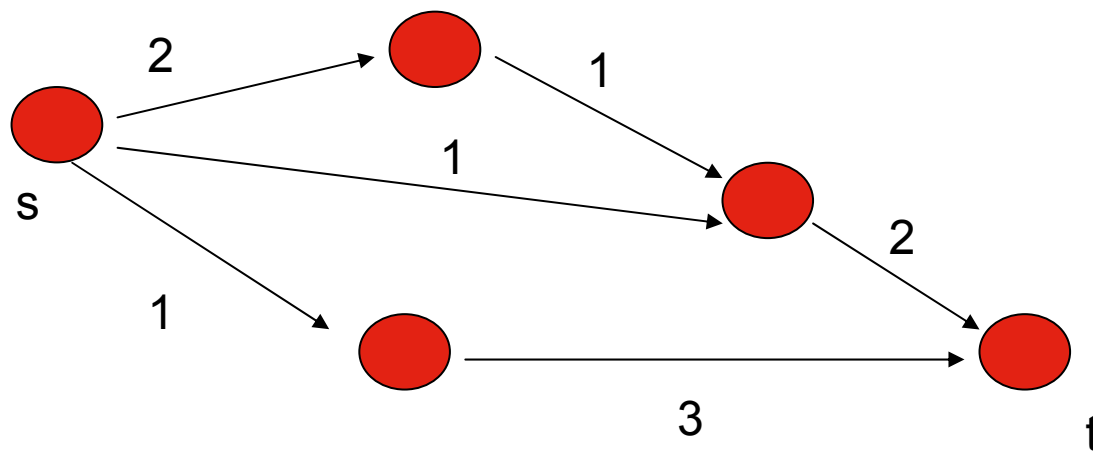
- Algorytm Dijkstry
- Algorytm Bellmana-Forda
- Algorytm Floyda-Warshalla
- Algorytm Johnsona





Algorytm Dijkstry - problem

- Najkrótsza ścieżka w grafie skierowanym, z wierzchołka s do wierzchołka t , w którym występują krawędzie o nieujemnych wagach





Algorytm Dikstrij - idea

- Pamiętany jest zbiór wierzchołków Q
- Nie obliczono dla nich najkrótszych ścieżek, wektor $D[i]$ odległości od wierzchołka s do i . Początkowo zbiór Q zawiera wszystkie wierzchołki a wektor D jest pierwszym wierszem macierzy wag krawędzi A .
Algorytm przebiega następująco:
- Dopóki zbiór Q nie jest pusty wykonuj:
- Pobierz ze zbioru Q wierzchołek v o najmniejszej wartości $D[v]$ i usuń go ze zbioru.
- Dla każdego następnika i wierzchołka v dokonaj relaksacji ścieżki, tzn. sprawdź, czy $D[i] > D[v] + A[v, i]$, tzn. czy aktualne oszacowanie odległości do wierzchołka i jest większe od oszacowania odległości do wierzchołka v plus waga krawędzi (v, i) . Jeżeli tak jest, to zaktualizuj oszacowanie $D[i]$ przypisując mu prawą stronę nierówności (czyli mniejszą wartość).



Algorytm Dijkstry - zastosowania

Algorytm ten jest używany w transmisji danych w sieciach komputerowych

W protokole OSPF jest wykorzystywany algorytm Dijkstry w celu znalezienia najlepszej drogi do przestania pakietów.





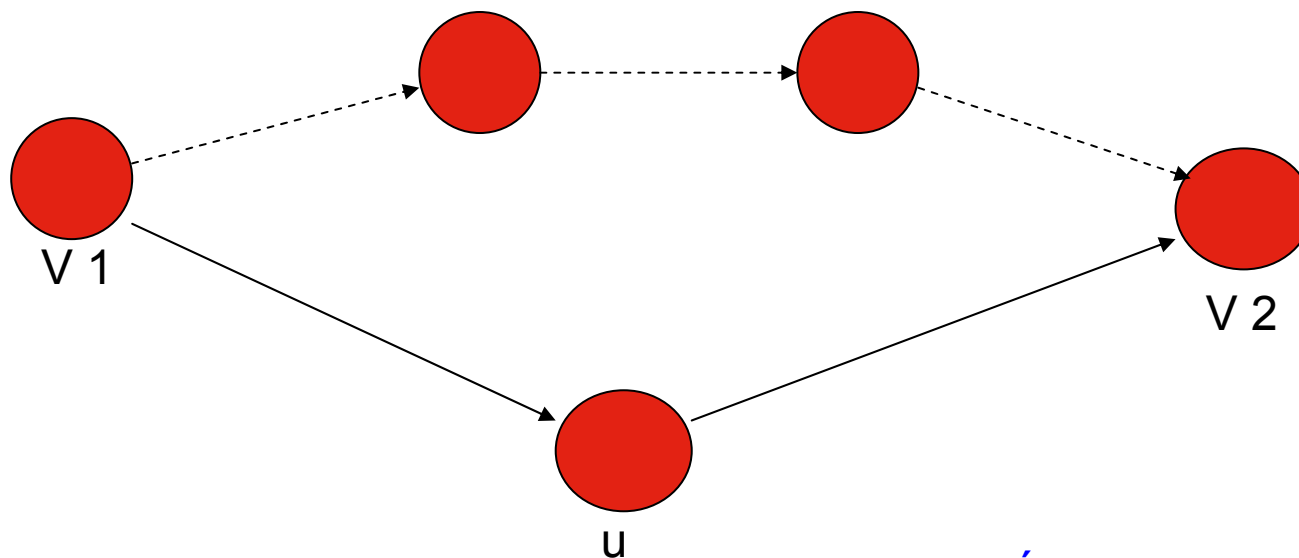
Algorytm Floyda-Warshalla - problem

- Większy zakres rozwiązywania problemu niż algorytm Dijkstry
- Służy do znajdowania najkrótszych ścieżek pomiędzy wszystkimi parami wierzchołków w grafie ważonym.



Algorytm Floyda-Warshalla - idea

- Algorytm Floyda-Warshalla korzysta, że jeśli najkrótsza ścieżka pomiędzy wierzchołkami $v1$ i $v2$ prowadzi przez wierzchołek u , to jest ona połączeniem najkrótszych ścieżek pomiędzy wierzchołkami $v1$ i u oraz u i $v2$.





Algorytm Floyda Warshalla- idea

- Działanie algorytmu zaczyna od utworzenia listy z długości najkrótszych ścieżek
- Następne kroki algorytmu włączają do przeliczeń kolejne wierzchołki. W konsekwencji n-ty krok algorytmu polega na sprawdzeniu, czy dla każdej pary wierzchołków nie da się skrócić obliczonej drogi.



Algorytm wyznaczania centralnego węzła w sieci – problem

- Poszukiwanie węzłów głównych
- Poszukiwanie węzłów znaczących
- Poszukiwanie węzłów centralnych



$$c_B(v) = \sum_{v \neq t \neq s} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

Algorytm wyznaczania centralności „betweenness” wężła w sieci – idea

- Istnieją wierzchołki $s, t, v \in V$
- Wiemy że istnieje przynajmniej jedna droga z s do t
- Liczymy ilość najkrótszych ścieżek z wierzchołka s do t ,
- Liczymy ilość ścieżek z s do t przechodzących przez wierzchołek v ,
- Suma stosunków σ_{st} i $\sigma_{st}(v)$ daje stopień centralności

$$\sigma_{st}$$

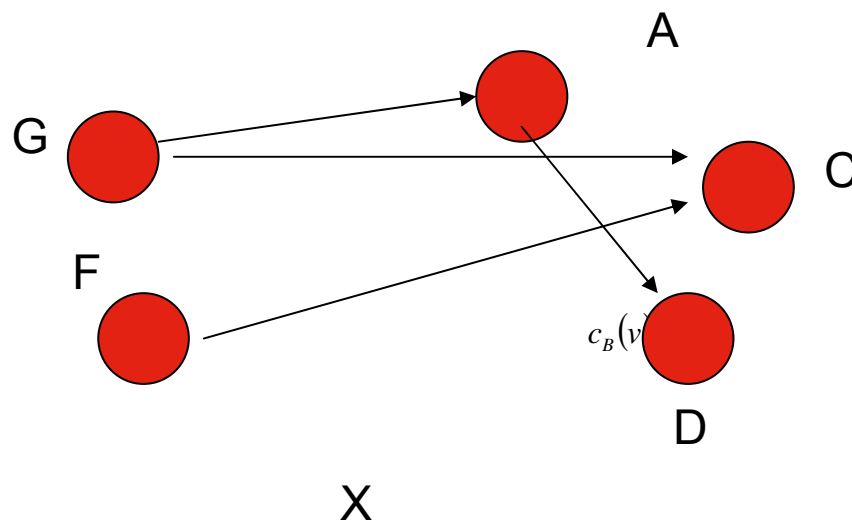
$$\sigma_{st}(v)$$

$$c_B(v) = \sum_{v \neq t \neq s} \frac{\sigma_{st}(v)}{\sigma_{st}}$$



Podejście do obliczania

- Ułatwienie obliczania polega na zbudowaniu tabelki z długościami i liczbą najkrótszych ścieżek pomiędzy wszystkimi parami wierzchołków.





Złożoność

- Obliczenia $c_B(v)$ wynoszą $O(n^2)$ tylko przy liczeniu wyznaczaniu centralności dla jednego wierzchołka
- Dla wszystkich wierzchołków złożoność tego algorytmu wynosi $O(n^3)$



Algorytm wyznaczania centralności wierzchołków według bliskości

- Policzenie wszystkich wierzchołków
- Policzenie sumy wszystkich odległości od wierzchołka v do wszystkich pozostałych wierzchołków
- Stosunek ilości wierzchołków ich odległości od wierzchołka v

$$n$$

$$\sum_{i=1, v \neq t}^{n-1} d(v, t_i)$$

$$c_c(v) = \frac{n-1}{\sum_{i=1, v \neq t}^{n-1} d(v, t_i)}$$



Środek ciężkości

- Środek ciężkości - parametr opisujący dany węzeł, obliczany na podstawie wszystkich najkrótszych ścieżek prowadzących do tego węzła.
- Współczynnik ma wysoką lub niską wartość. Im wyższa wartość tym wierzchołek leży bardziej w centrum sieci
- Węzeł reprezentujący osobę o największym środku ciężkości można rozumieć jako osobę, która posiada bardzo dużo informacji.



Stopnie wierzchołków przychodzących i wychodzących

- Stopnie wierzchołków przychodzących i wychodzących
- Miara jest obliczana na podstawie ilości krawędzi wchodzących do danego wężła oraz z niego wychodzących.
- Parametr dla danego wężła jest tym większy, im więcej ma on połączeń przychodzących lub wychodzących.



Algorytmy wyznaczania centralnego węzła w sieci – zastosowanie

- Wskazanie węzła o największej centralności
- Możliwość analizy połączeń bezpośrednich
- Możliwość badania pozycji w grupie
- Analiza straty



Problemy z liczeniem centralności

- Analiza dużej ilości sieci w racjonalnym czasie
- Złożoność obliczeniowa algorytmów wyznaczających centralność w sieci - większość algorytmów rozwiązuje problemy w wielomianowym czasie



Zastosowanie aproksymacji

- Aproksymacja daje rozsądne lecz przybliżone rozwiązania
- Stosując aproksymację wybiera się kompromis pomiędzy czasem pracy - analizą sieci przez program, a dokładnością obliczeń



Przybliżenie rozwiązań

- Przybliżenie powinno dotyczyć przede wszystkim sposobu rozwiązywania problemów oraz w efekcie powinno się skupić na jakości wyników , które daje algorytmy przybliżone



Źródła do prezentacji

- „Algorithms Centrality Indices” D. Koschützki, K. Lehmann, L. Peeters, S. Richter, D. Tenfelde-Podehl and O. Zlotowski
- „Jonak 07_Sieci_spoleczne_i_dynamika_sieciowa” A. Jonak
- „Środowisko do identyfikowania” wzorców zachowań w oparciu o podejście sieci społecznych” W. Piekaj¹, G. Skorek¹, A. Zygmunt¹, J. Kozlak¹
- „Badanie cytowania FN” J. Krzysztofiak Szopa
- <http://www.algorytm.org/>



Koniec

- Dziękuję za uwagę