



Politechnika Wroclawska

Centralność w sieciach społecznych

Radostaw Michalski
Social Network Group - kwiecień 2009



Agenda spotkania

- Pojęcie centralności
- Potrzeba pomiaru centralności
- Miary centralności
 - degree centrality
 - betweenness
 - closeness
 - Eigenvector (wektory własne)
- Problemy



Czym jest centralność

- Istotność członka sieci społecznej
- Mierzona różnymi miarami, w zależności od celu
- Określenie centralności w sieciach społecznych pozwala wyłonić kluczowych użytkowników
- Centralność może dotyczyć węzłów jak i całej sieci



Potrzeba pomiaru centralności

- Analiza węzłów kluczowych
 - zatrzymanie użytkownika w sieci
 - potencjalnie duży oddźwięk w przypadku zwiększenia nakładów w tym miejscu
 - porównanie istotności stron
- Analiza sieci
 - w przypadku sieci zdecentralizowanej brak „single point of failure”
 - w przypadku sieci scentralizowanej - potrzeba wprowadzenia ścieżek alternatywnych (zależy od rodzaju sieci)



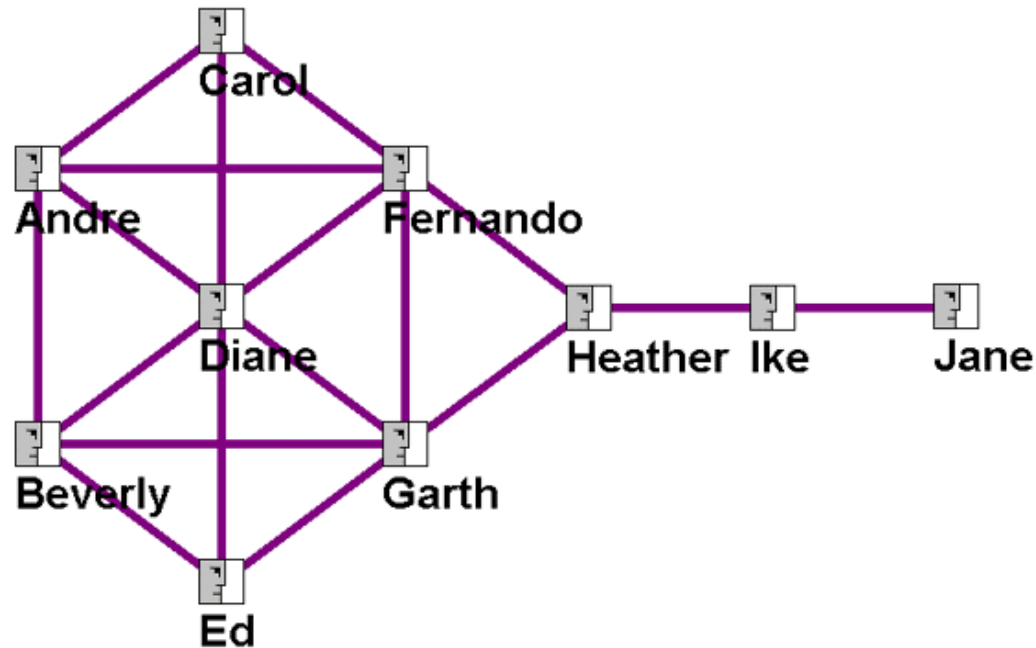
Miary centralności

- Degree centrality
 - ogólnie
 - outdegree centrality (DC)
 - indegree centrality (prestige)
- Betweenness centrality
- Closeness centrality
- Eigenvector



Degree centrality - ogólnie

- Ilość połączeń danego użytkownika z innymi w grafach nieskierowanych



$$C_D(v) = \frac{\deg(v)}{n - 1}$$

- Na powyższym rysunku - Diane



Outdegree centrality

- Dotyczy grafów skierowanych
- Węzeł jest bardziej istotny gdy komunikuje się z większą ilością użytkowników
- Im więcej krawędzi wychodzących, tym „lepiej”
- Przykładowa interpretacja: im większa wartość, tym bardziej towarzyska jest dana osoba



Indegree centrality (prestige)

- Dotyczy grafów skierowanych
- Węzeł jest bardziej istotny gdy wskazuje na niego większa ilość innych wierzchołków
- Im więcej wierzchołków wchodzących, tym „lepiej”
- Przykładowa interpretacja: im większa wartość, tym bardziej popularna jest dana osoba



Betweenness

- Kluczowość wierzchołka w zakresie komunikacji (przechodniość, pośredniczenie)
- Możliwość wykrycia pojedynczych miejsc awarii
- Wysoka wartość - możliwa interpretacja jako punkt krytyczny dla sieci



Algorytm wyznaczania centralności „betweennes” węzła 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

$$\sigma_{st}$$

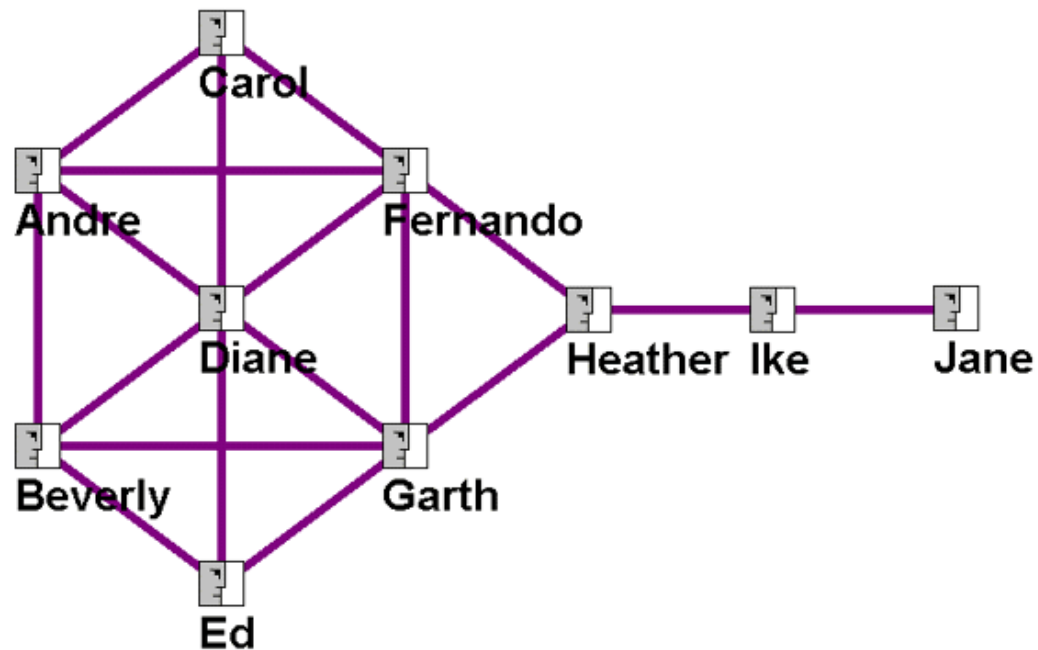
$$\sigma_{st}(v)$$

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



Betweenness - przykład

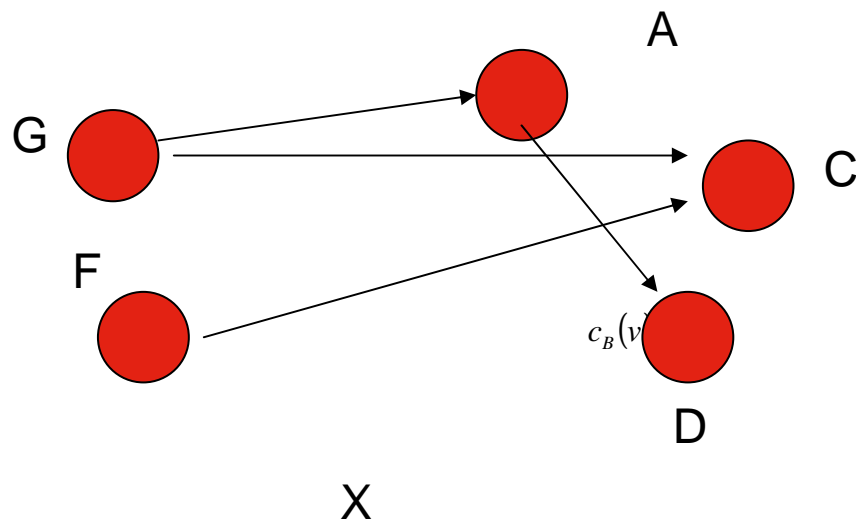
- Na poniższym przykładzie - Heather





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 $O(n^3)$



Closeness centrality

- Stopień bliskości
- Określa jak daleko / jak blisko wierzchołek ma do pozostałych wierzchołków w sieci
- Wysoki stopień centralności wskazuje na użytkowników, którzy mogą mieć dobre własności propagacji informacji (wirusów itp. także...)



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

$$n$$

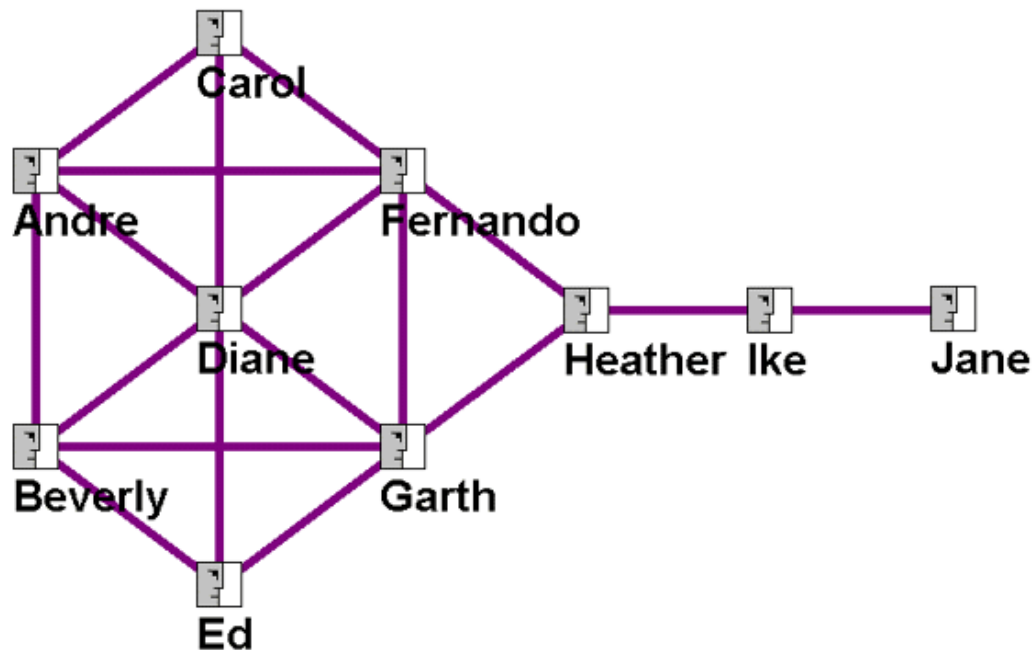
$$\sum_{i=1, v \neq t}^{n-1} d(v, t_i)$$

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



Closeness - przykład

- Na poniższym przykładzie - Fernando i Garth





Eigenvector (wektory własne)

- Przykład web centrality - która strona w sieci na zadany temat jest najistotniejsza?
- Nie wystarczy ilość wskazań do tej strony, ale ważne skąd te wskazania pochodzą (wskazujący też muszą być istotni, inaczej łatwo oszukać engine)
- Eigenvector do obliczenia pozycji strony uwzględnia pozycje stron pozostałych
- Przykłady: PageRank, Social Position



Problemy

- Duża złożoność obliczeniowa
 - aproksymacja?
- Duża dynamika sieci często kłopotliwa (wymaga ponownych obliczeń w algorytmach liczących całą sieć)
- Konieczność doboru właściwej miary i wyciągnięcia właściwych wniosków
- Spoofowanie algorytmów - np. PageRank



Politechnika Wroclawska

Pytania?



Źródła

- Musial, K., Juszczyzyn, K., Gabrys, B. and Kazienko, P., 2008. „Patterns of Interactions in Complex Social Networks Based on Coloured Motifs Analysis”, 15th International Conference on Neural Information Processing (ICONIP'2008), 25-28 Nov 2008, Auckland, New Zealand,
- Wojciech Piekaj, Grzegorz Skorek, Anna Zygmunt, Jarosław Kozlak „Środowisko do identyfikowania wzorców zachowań w oparciu o podejście sieci społecznych”
- „Algorytmy wyznaczania centralności w sieci”, Szymon Szyłko
- Social Network Analysis - <http://www.orgnet.com/sna.html>