

POLITECHNIKA WROCŁAWSKA
WYDZIAŁ INFORMATYKI I ZARZĄDZANIA

KIERUNEK: INFORMATYKA

PROJEKT
Teoria i inżynieria ruchu teleinformatycznego

Analiza sieci systemów autonomicznych przy
użyciu algorytmów grafowych

AUTORZY:

Bartosz Cieśla, Bartosz Janusz, Bartosz Kardas

WROCŁAW, 2017

Spis treści

1. Wstęp	5
1.1. Cel projektu	5
1.2. Dane sieci	5
1.3. Systemy autonomiczne	5
2. Algorytmy	6
2.1. Centralność w grafach	6
2.2. Betweenness Centrality	6
2.2.1. Algorytm wyznaczania	6
2.3. Closeness Centrality	7
2.4. Eigenvector Centrality - Pagerank	8

Spis rysunków

2.1. Działanie Betweenness Centrality na przykładowym grafie	7
2.2. Działanie Closeness Centrality na przykładowym grafie	8
2.3. Działanie Eigenvector Centrality na przykładowym grafie	9
2.4. Działanie Betweenness Centrality na przykładowym grafie	10
2.5. Działanie Betweenness Centrality na przykładowym grafie	10
2.6. Działanie Betweenness Centrality na przykładowym grafie	11
2.7. Działanie Betweenness Centrality na przykładowym grafie	11
2.8. Działanie Betweenness Centrality na przykładowym grafie	12
2.9. Działanie Betweenness Centrality na przykładowym grafie	12
2.10. Działanie Betweenness Centrality na przykładowym grafie	13
2.11. Działanie Betweenness Centrality na przykładowym grafie	13
2.12. Działanie Betweenness Centrality na przykładowym grafie	14
2.13. Działanie Betweenness Centrality na przykładowym grafie	14
2.14. Działanie Betweenness Centrality na przykładowym grafie	15
2.15. Działanie Betweenness Centrality na przykładowym grafie	16
2.16. Działanie Betweenness Centrality na przykładowym grafie	17
2.17. Działanie Betweenness Centrality na przykładowym grafie	17
2.18. Działanie Betweenness Centrality na przykładowym grafie	18
2.19. Działanie Betweenness Centrality na przykładowym grafie	18
2.20. Działanie Betweenness Centrality na przykładowym grafie	19
2.21. Działanie Betweenness Centrality na przykładowym grafie	20
2.22. Działanie Betweenness Centrality na przykładowym grafie	21
2.23. Działanie Betweenness Centrality na przykładowym grafie	21
2.24. Działanie Betweenness Centrality na przykładowym grafie	22
2.25. Działanie Betweenness Centrality na przykładowym grafie	22
2.26. Działanie Betweenness Centrality na przykładowym grafie	23
2.27. Działanie Betweenness Centrality na przykładowym grafie	24
2.28. Działanie Betweenness Centrality na przykładowym grafie	25
2.29. Działanie Betweenness Centrality na przykładowym grafie	25
2.30. Działanie Betweenness Centrality na przykładowym grafie	26
2.31. Działanie Betweenness Centrality na przykładowym grafie	26

Spis tabel

Rozdział 1

Wstęp

1.1. Cel projektu

Celem projektu jest ukazanie oraz dokonanie pomiarów zmienności rzeczywistych sieci systemów autonomicznych. Sieci te są reprezentowane za pomocą grafów nieskierowanych. Do ich analizy zastosowano algorytmy typowo wykorzystywane w przypadku badania sieci społecznościowych, tak zwane miary centralności grafu. Oprócz nich szczególną uwagę zwrócono na stopień poszczególnych węzłów.

1.2. Systemy autonomiczne

Oryginalnie projekt miał opierać się na analizie grafów, gdzie każdy wierzchołek odpowiadał pojedynczemu urządzeniu - routerowi. Ze względu jednak na niską dostępność takich zestawów danych oraz ich gigantyczny rozmiar (niemożliwy do przetworzenia na dostępnym sprzęcie), zdecydowano się na badanie większych jednostek. System autonomiczny to sieć lub grupa sieci opartych na protokole IP pod wspólną administracyjną kontrolą, w której utrzymywany jest spójny schemat trasowania. Struktury te są podstawową jednostką budulcową Internetu na poziomie domen. Większość dostępnych topologii sieci internetowej opiera się na ich strukturze, dlatego uznano, że i w opisywanym projekcie znajdą zastosowanie.

1.3. Dane sieci

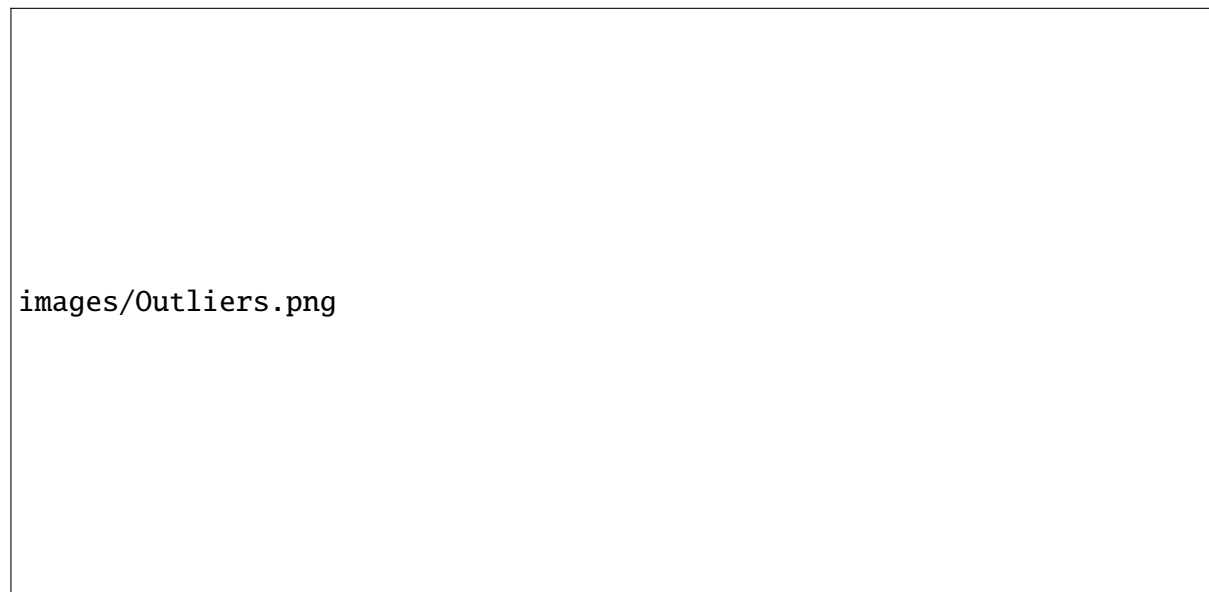
Znalezienie odpowiednich danych wejściowych wymagało trochę czasu. Z racji niedeterministycznej struktury badanych sieci, niemożliwe było wygenerowanie danych testowych. Konieczne okazało się wykorzystanie danych pochodzących z badań istniejącej sieci internetu. Zdecydowano się wykorzystać zestaw pochodzący z projektu University of Oregon Route Views. Zawiera on 733 instancje grafu, tworzonego poprzez codzienne badania struktury sieci na poziomie systemów autonomicznych. Najstarsza instancja pochodzi z 8 listopada 1997 roku, najmłodsza z 2 stycznia 2000 roku. Dane były zbierane za pomocą techniki multi-hop z wykorzystaniem protokołu BGP.

1.4. Wstępne przygotowanie danych

Dane zawarte w zestawie mają postać listy krawędzi. Mimo iż są to grafy nieskierowane, wszystkie krawędzie występują w postaci zdublowanej oraz część węzłów posiada pętle. Warto tu nadmienić, że taki zapis grafu w pewnym kontekście na pewno miał określone znaczenie. Jednakże

z punktu widzenia zamierzonych badań oba zjawiska były niepożądane, w związku z czym napisano własną funkcję wczytującą graf do programu. Jej zadaniem było odfiltrowanie krawędzi wielokrotnych oraz pętli.

Drugim problemem na który natknęto się po wczytaniu grafów, były duże błędy pomiarowe. Po stworzeniu wykresu liczby wierzchołków dla wszystkich instancji zauważono, że istotna ich część znacząco odstaje od widocznego trendu (w kierunku zera). Wykres liczby krawędzi ukazał spadki w tych samych miejscach, co utwierdziło autorów o błędach w pomiarach. Podejrzewają oni, że było to spowodowane przerwaniem procesu zbierania danych o strukturze sieci. Zaskutkowało to mniejszymi instancjami grafów dla niektórych dni. W celu ujednolicenia danych zastosowano funkcję usuwającą obserwacje odstające zawartą w oprogramowaniu Matlab *isoutlier()*. Efekt jej działania można zobaczyć poniżej.



Rys. 1.1: Wykres liczby wierzchołków dla poszczególnych instancji z zaznaczonymi obserwacjami odstającymi

Rozdział 2

Algorytmy

2.1. Centralność w grafach

W teorii grafów wskaźniki centralności informują o najbardziej znaczących wierzchołkach grafu. Ich przykładowymi zastosowaniami mogą być: znalezienie lidera, przywódcy spośród danej grupy osób, ustalenie kluczowego elementu infrastruktury sieciowej lub miejskiej bądź znalezienie osobnika o największym potencjale do roznoszenia choroby. Istnieje wiele odmiennych wskaźników centralności. Zrealizowany projekt implementuje trzy z nich: Closeness Centrality, Betweenness Centrality oraz Pagerank (jedna z odmian Eigenvector Centrality)

2.2. Betweenness Centrality

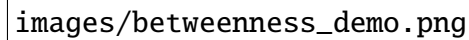
Określa kluczowość wierzchołka w zakresie komunikacji - przechodność, pośredniczenie. Czyli w jakim stopniu dany wierzchołek jest spoiwem dla danej sieci. Jest to miara o bardzo wielkiej wartości, gdyż dzięki niej można znaleźć punkty krytycznej sieci bądź grafu.

2.2.1. Algorytm wyznaczania

1. Wyznaczyć ilość najkrótszych ścieżek między wierzchołkami u i v (d_{uv})
2. Wyznaczyć ilość najkrótszych ścieżek między wierzchołkami u i v , które przechodzą przez wierzchołek w ($d_{uv}(w)$)
3. Suma stosunków oznacza stopień centralności wierzchołka w

$$c_b(w) = \sum_{u \neq v \neq w} \frac{d_{uv}(w)}{d_{uv}}$$

Przykład



images/betweenness_demo.png

Rys. 2.1: Działanie Betweenness Centrality na przykładowym grafie

2.3. Closeness Centrality

Jest to stopień bliskości. Określa jak blisko (daleko) wierzchołek ma do pozostałych w grafie. Wysoki stopień bliskości świadczy o dobrej własności propagacji informacji w grafie - element ten szybko rozprowadzi daną wiadomość (wirusa itp) po całej sieci.

Algorytm wyznaczania

1. Wyznaczyć odległości pomiędzy wierzchołkiem u a pozostałymi wierzchołkami w grafie v (d_{uv})
2. W zależności od rodzaju grafu zsumować otrzymane odległości:
 1. Dla grafów rzadkich

$$c_c(u) = \frac{1}{\sum d_{uv}}$$

2. Dla grafów silnie połączonych

$$c_c(u) = \sum_{u \neq v} \frac{1}{d_{uv}}$$

Przykład



Rys. 2.2: Działanie Closeness Centrality na przykładowym grafie

2.4. Eigenvector Centrality - Pagerank

Określa wpływ, oddziaływanie wierzchołka na pozostałe w grafie. Wykorzystuje nie tylko ilość połączeń danego wierzchołka z innymi, a przede wszystkim ich jakość. Wartości przypisane do każdego z wierzchołków bazują na koncepcji w której wysoko ocenione wierzchołki bardziej wpływają na ostateczną ocenę połączonego wierzchołka, niż te, których ocena jest niska. Jedną z odmian Eigenvector Centrality jest algorytm PageRank. Poniżej przedstawiono uproszczony algorytm jego działania.

Algorytm wyznaczania

1. Wyznaczyć ilość wierzchołków w grafie (N)
2. Wyznaczyć stopień każdego z wierzchołków ($l(u)$)
3. Zainicjować wartości początkowe dla każdego wierzchołka wartością początkową ($c_e(u) = 1$)
4. Określić współczynnik tłumienia, zwykle wynosi on około 0.85 ($d = 0.85$)
5. Obliczyć nową wartość PageRank każdego wierzchołka

$$c_e(u) = \frac{1-d}{N} + d \sum_{v \in B_u} \frac{c_e(v)}{l(v)}$$

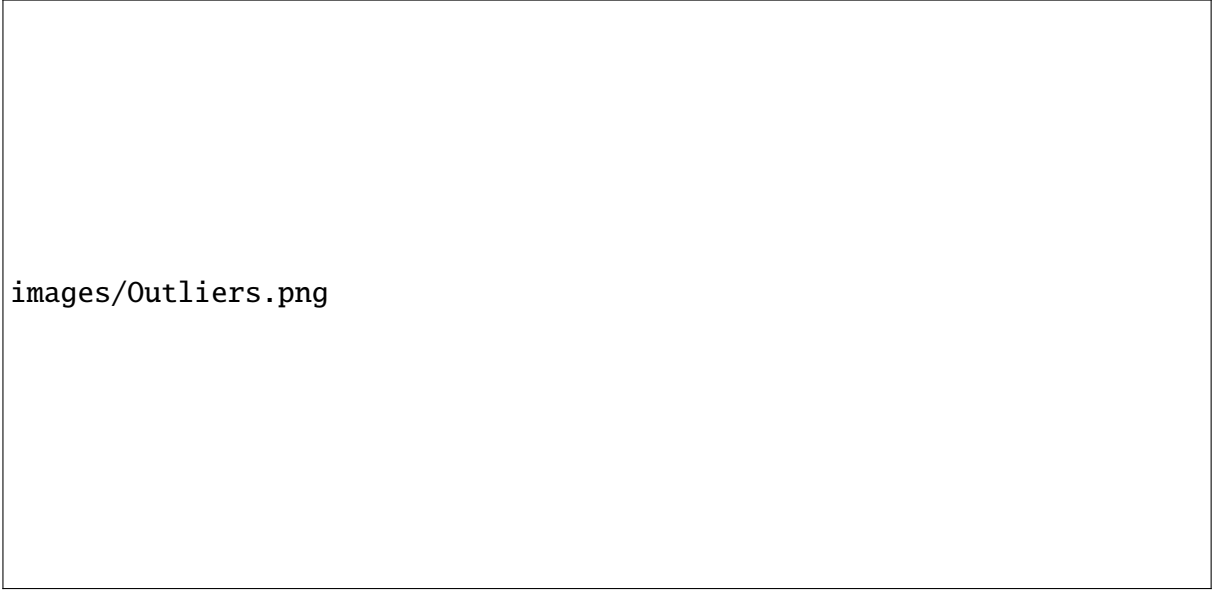
B_u oznacza zbiór wszystkich wierzchołków, które odnoszą się do wierzchołka u

Przykład



Rys. 2.3: Działanie Eigenvector Centrality na przykładowym grafie

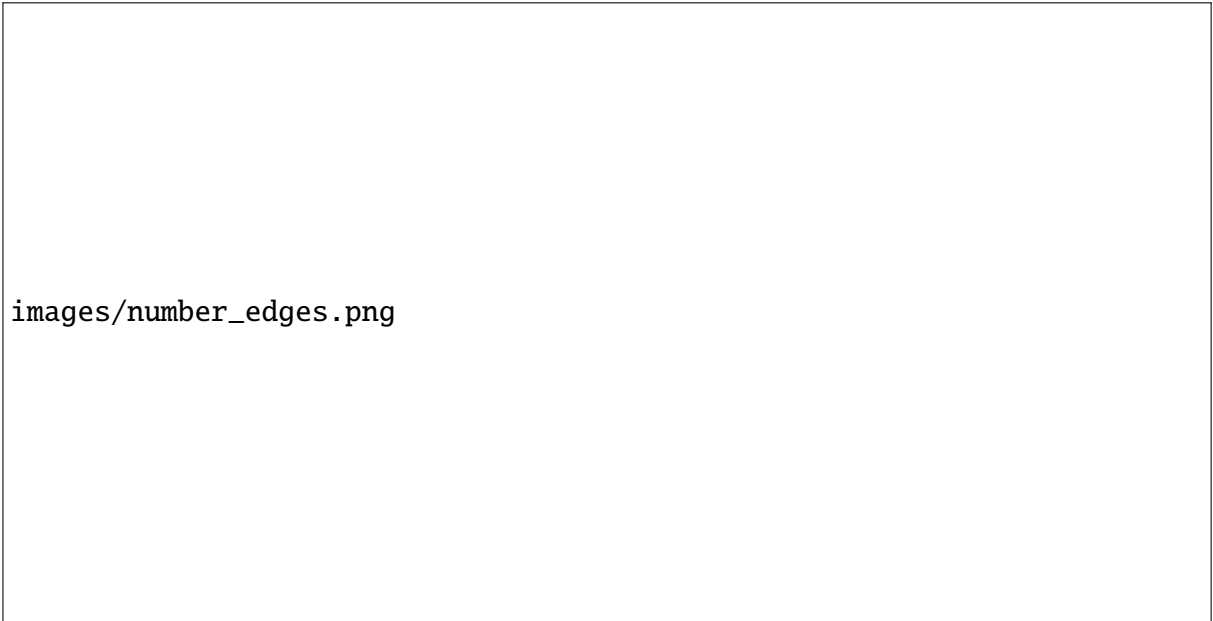
Przykład



images/Outliers.png

Rys. 2.4: Działanie Betweenness Centrality na przykładowym grafie

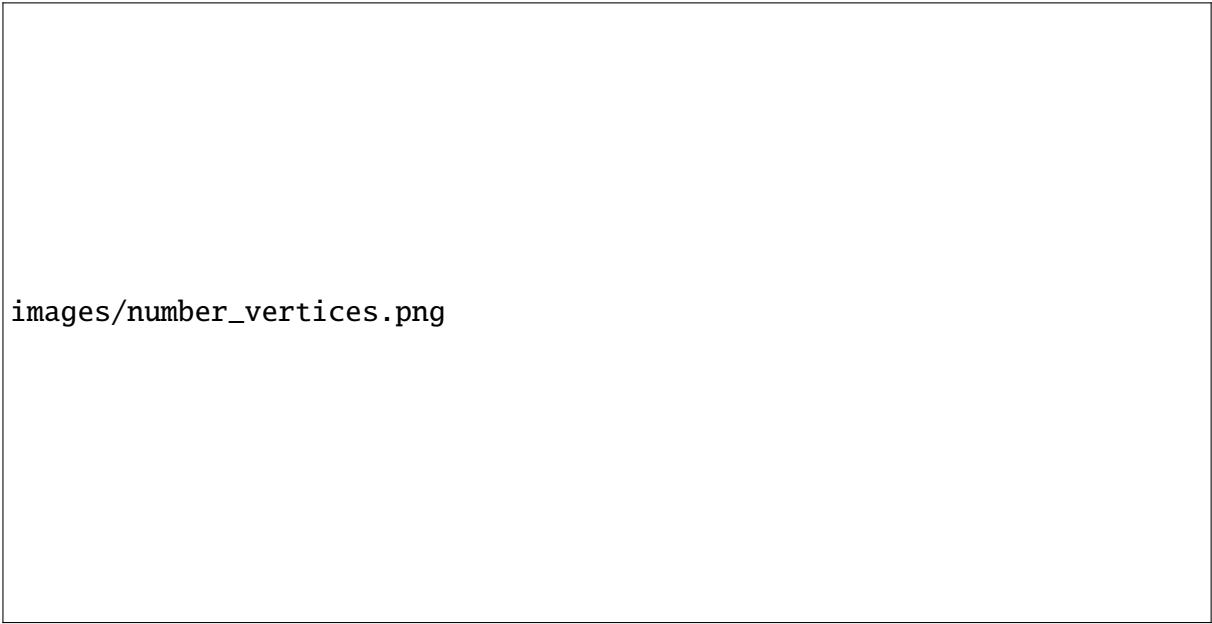
Przykład



images/number_edges.png

Rys. 2.5: Działanie Betweenness Centrality na przykładowym grafie


Przykład



images/number_vertices.png

Rys. 2.6: Działanie Betweenness Centrality na przykładowym grafie

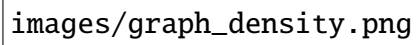
Przykład



images/degree_percentiles.png

Rys. 2.7: Działanie Betweenness Centrality na przykładowym grafie

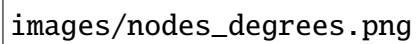
Przykład



images/graph_density.png

Rys. 2.8: Działanie Betweenness Centrality na przykładowym grafie

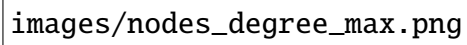
Przykład



images/nodes_degrees.png

Rys. 2.9: Działanie Betweenness Centrality na przykładowym grafie

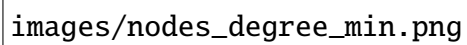
Przykład



images/nodes_degree_max.png

Rys. 2.10: Działanie Betweenness Centrality na przykładowym grafie


Przykład



images/nodes_degree_min.png

Rys. 2.11: Działanie Betweenness Centrality na przykładowym grafie

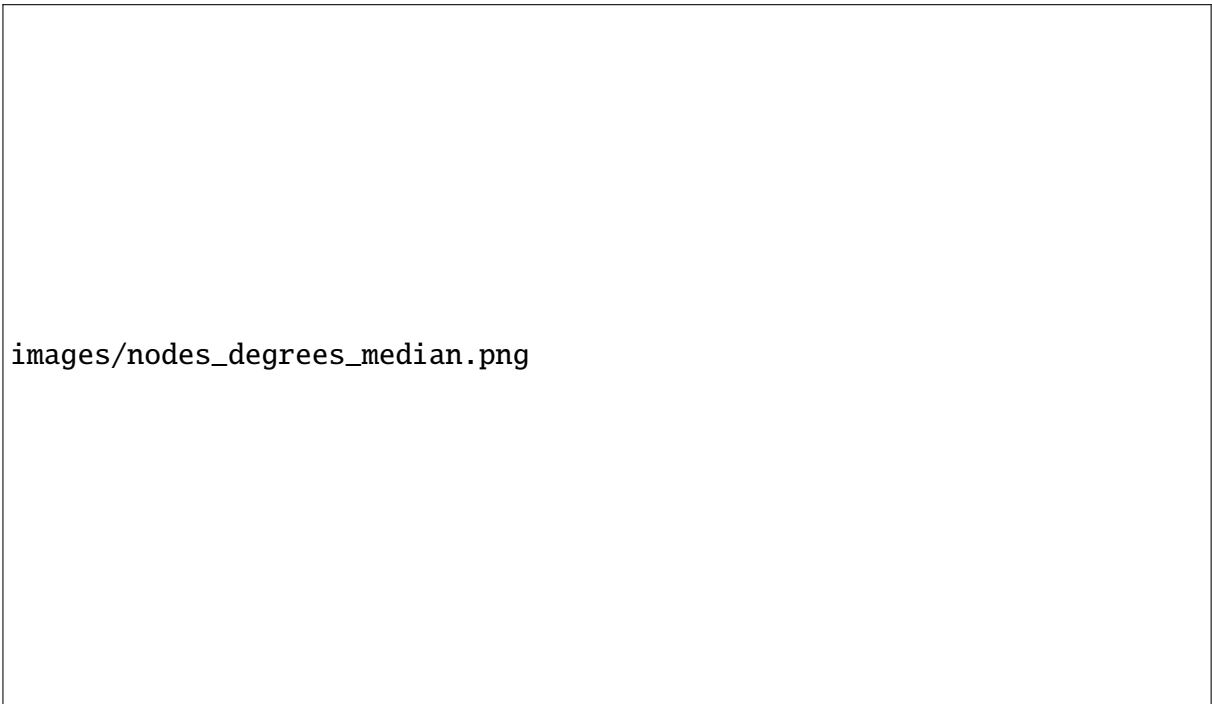
Przykład



images/nodes_degrees_mean.png

Rys. 2.12: Działanie Betweenness Centrality na przykładowym grafie

Przykład



images/nodes_degrees_median.png

Rys. 2.13: Działanie Betweenness Centrality na przykładowym grafie

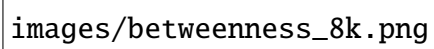
Przykład



images/betweenness.png

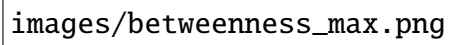
Rys. 2.14: Działanie Betweenness Centrality na przykładowym grafie

Przykład

A large rectangular box with a thin black border. Inside the box, on the left side, is the text "images/betweenness_8k.png". The rest of the box is empty.

Rys. 2.15: Działanie Betweenness Centrality na przykładowym grafie

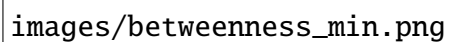
Przykład



images/betweenness_max.png

Rys. 2.16: Działanie Betweenness Centrality na przykładowym grafie

Przykład



images/betweenness_min.png

Rys. 2.17: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/betweenness_mean.png

Rys. 2.18: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/betweenness_median.png

Rys. 2.19: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/closeness.png

Rys. 2.20: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/closeness_8k.png

Rys. 2.21: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/closeness_max.png

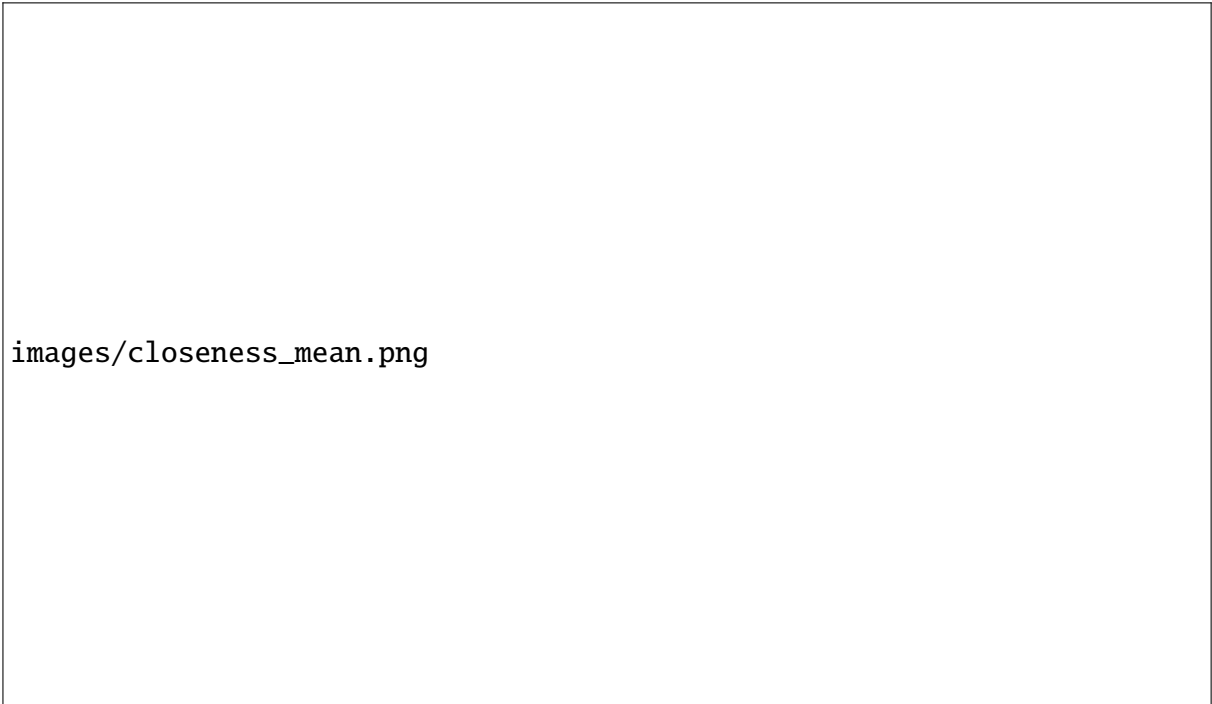
Rys. 2.22: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/closeness_min.png

Rys. 2.23: Działanie Betweenness Centrality na przykładowym grafie

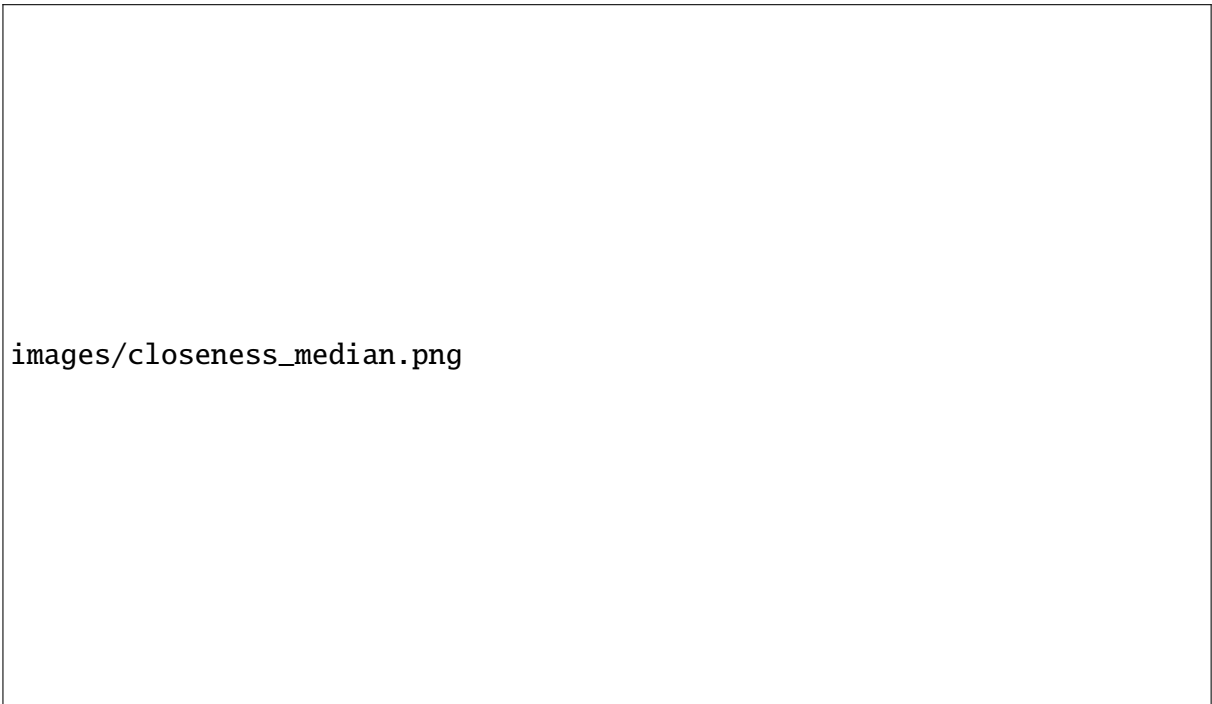
Przykład



images/closeness_mean.png

Rys. 2.24: Działanie Betweenness Centrality na przykładowym grafie

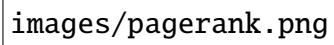
Przykład



images/closeness_median.png

Rys. 2.25: Działanie Betweenness Centrality na przykładowym grafie

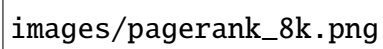
Przykład



images/pagerank.png

Rys. 2.26: Działanie Betweenness Centrality na przykładowym grafie

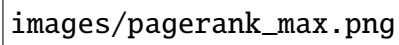
Przykład



images/pagerank_8k.png

Rys. 2.27: Działanie Betweenness Centrality na przykładowym grafie

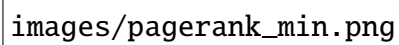
Przykład



images/pagerank_max.png

Rys. 2.28: Działanie Betweenness Centrality na przykładowym grafie

Przykład



images/pagerank_min.png

Rys. 2.29: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/pagerank_mean.png

Rys. 2.30: Działanie Betweenness Centrality na przykładowym grafie

Przykład

images/pagerank_median.png

Rys. 2.31: Działanie Betweenness Centrality na przykładowym grafie