

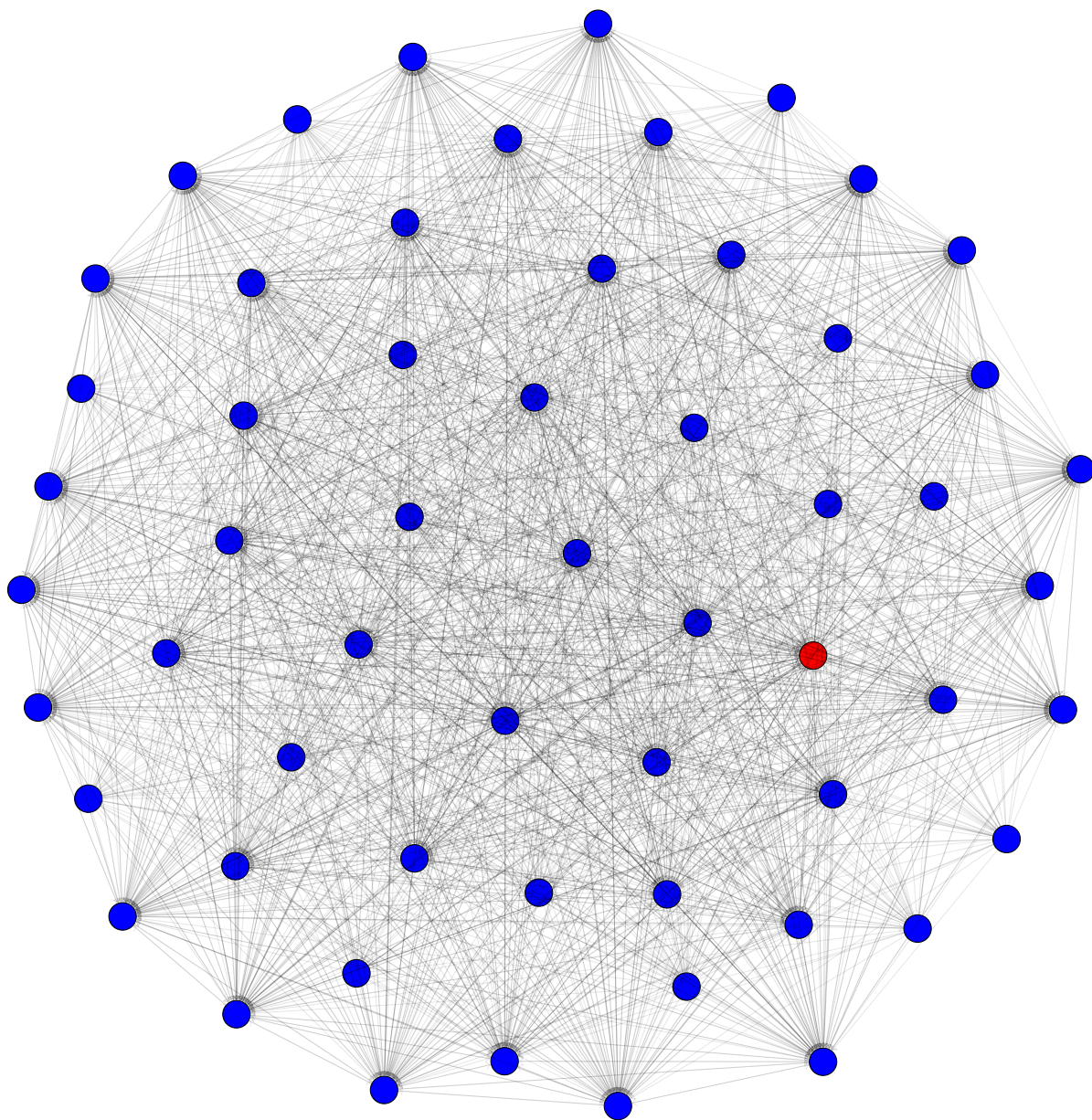
## Modelowanie sieci

# Robot internetowy i graf WWW

Mateusz Stencel

188676

---



Wizualizacja grafu wybranej domeny dla 50 stron.

# Spis treści

<b>1. Robot internetowy</b>	<b>3</b>
1.1. Wstęp . . . . .	3
1.2. Zasada działania robota . . . . .	3
1.3. Implementacja . . . . .	3
1.4. Czas działania . . . . .	4
<b>2. Analiza grafu WWW</b>	<b>5</b>
2.1. Parametry ogólne . . . . .	5
2.2. Składowe spójności . . . . .	5
2.3. Rozkłady stopni . . . . .	6
2.4. Ścieżki . . . . .	7
2.5. Współczynniki klasteryzacji . . . . .	8
2.6. Spójność wierzchołkowa . . . . .	9
2.7. Pagerank . . . . .	10
<b>3. Odporność na ataki</b>	<b>12</b>
3.1. Awarie stron . . . . .	12
3.1.1. Rozkłady stopni . . . . .	12
3.1.2. Składowe spójności . . . . .	14
3.1.3. Odległości . . . . .	14
3.2. Ataki . . . . .	16
3.2.1. Rozkłady stopni . . . . .	16
3.2.2. Składowe spójności . . . . .	17
3.2.3. Odległości . . . . .	18
3.3. Wnioski . . . . .	19
<b>Spis rysunków</b>	<b>20</b>
<b>Spis tabel</b>	<b>21</b>

# 1. Robot internetowy

## 1.1 Wstęp

Celem projektu było stworzenie robota internetowego, który przeszukuje zasoby w obrębie domeny internetowej. Zadaniem robota było pobieranie i zapisywanie na dysku kopii dokumentów oraz jednocześnie budowanie grafu połączeń między nimi.

W ramach realizacji projektu, zbadano domenę zagranicznej uczelni [The University of Manchester](#), której zasoby zostały przeszukane przez zaprojektowanego robota. Projekt został zaimplementowany w języku **Python** z wykorzystaniem biblioteki **networkx** do stworzenia i analizowania grafu połączeń między stronami. Podczas projektu przyjęto ograniczenie do 3500 stron.

## 1.2 Zasada działania robota

Robot internetowy to program komputerowy, który automatycznie przeszukuje strony internetowe w celu zbierania danych i indeksowania zawartości stron. Roboty są podstawą dla wyszukiwarek internetowych, takiej jak Google. Dzięki robotom internetowym możliwe jest szybkie gromadzenie danych z różnych źródeł w sieci oraz ich analiza

Swoją pracę robot opiera na przestrzeganiu Robots Exclusion Protocol (REP), który jest standardem określającym zasady dotyczące dostępu robotów internetowych do zasobów witryny. REP informuje, które strony lub zasoby mogą być przeszukiwane, a które powinny zostać pominięte. opiera się na pliku tekstowym o nazwie `robots.txt`, który jest umieszczany w głównym katalogu witryny i może wyglądać następująco:

```
User-agent: *  
Disallow: /private/  
Allow: /public/
```

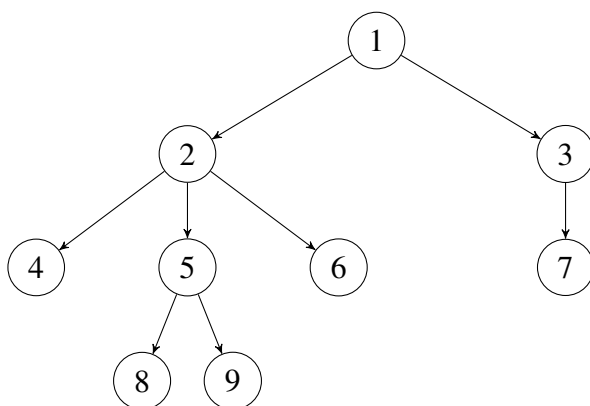
W tym przykładzie:

- `User-agent: *` oznacza, że zasady dotyczą wszystkich robotów,
- `Disallow: /private/` zabrania robotom odwiedzania strony w katalogu `private`,
- `Allow: /public/` zezwala robotom na dostęp do strony w katalogu `public`.

## 1.3 Implementacja

Robot internetowy został zaimplementowany w oparciu o algorytm przeszukiwania grafu w szerz (BFS, *ang. Breadth-First Search*). Podczas działania robot przeszukuje zasoby domeny

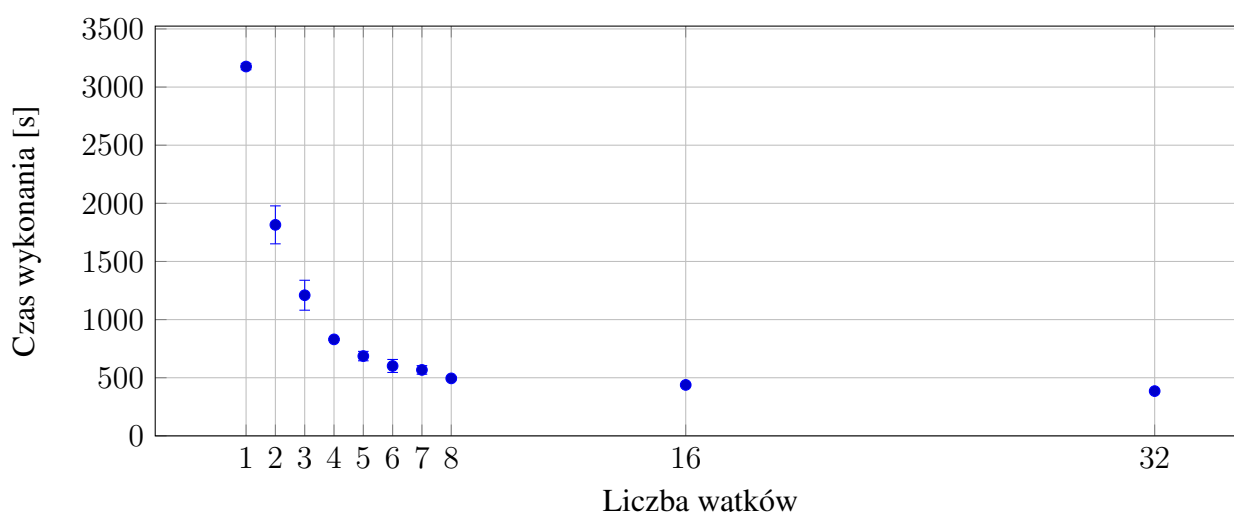
zaczynając od strony początkowej, a następnie odwiedza kolejne strony, do których prowadzą znalezione odnośniki. Nowe strony są dodawane do kolejki, a robot przetwarza je w kolejności ich odkrycia. Sposób przechodzenia pokazuje Rys. 1.1.



Rysunek 1.1: Przejście po grafie wg. algorytmu BFS.

## 1.4 Czas działania

W celu zbadania wpływu liczby wątków na czas działania robota internetowego, wykonano serię pomiarów, których wyniki przedstawiono na Rys. 1.2. Na podstawie wyników można zaobserwować, że zwiększanie liczby wątków prowadzi do istotnego skrócenia czasu działania programu. Powyżej czterech przyrost wydajności jest już mniej wyraźny, który może związany z narzutem przełączania wątków, a także kosztem synchronizacji jakim jest dodawanie elementów do grafu sieci. Błędy pomiarowe są niewielkie, a ich obecność może wynikać z jakości połączenia lub stanu zajętości samej domeny.



Rysunek 1.2: Czas wykonania programu w zależności od liczby wątków.

## 2. Analiza grafu WWW

### 2.1 Parametry ogólne

Graf prosty  $G$  składa się z niepustego zbioru skończonego  $V(G)$ , którego elementy nazywane są wierzchołkami, i skończonego zbioru  $E(G)$  różnych par nieuporządkowanych różnych elementów  $V(G)$ , którego nazywane są krawędziami. Badany graf jest digrafem, co oznacza, że jego krawędzie są skierowane. Na rysunku 2.1 pokazano przykładowe rozróżnienie między grafem prostym a digrafem.



Rysunek 2.1: Porównanie grafu prostego i digrafu.

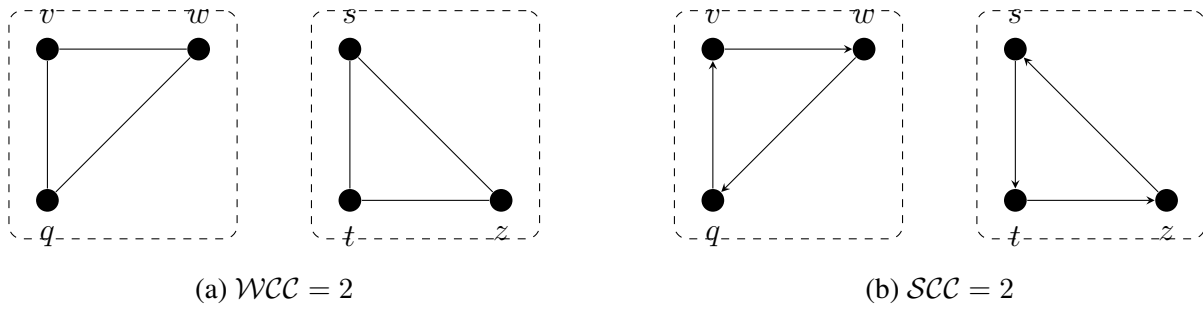
W tabeli 2.1 pokazano liczbę wierzchołków i łuków dla grafu www. Liczba łuków wskazuje, że graf jest dosyć gęsty.

Parametr	Wartość
Liczba wierzchołków	3500
Liczba łuków	431129

Tabela 2.1: Podstawowe parametry badanego grafu.

### 2.2 Składowe spójności

W grafach skierowanych wyróżnia się różne rodzaje składowych spójności. Silna składowa spójności – *SCC* (ang. *Strongly Connected Component*) to maksymalny podzbiór wierzchołków, w którym dla każdej pary wierzchołków  $u$  oraz  $v$  istnieje ścieżka z  $u$  do  $v$  oraz ścieżka z  $v$  do  $u$  w skierowanym grafie. Słaba składowa spójności – *WCC* (ang. *Weakly Connected Component*) definiowana jest jako maksymalny podzbiór wierzchołków, w którym każda para wierzchołków jest połączona ścieżką po usunięciu kierunków krawędzi. Przykładowe rozróżnienie na *SCC* i *WCC* pokazano na Rys. 2.2.



Rysunek 2.2: Podział na  $WCC$  i  $SCC$ .

W analizie struktury grafu skierowanego wyróżnia się również komponenty typu  $IN$  oraz  $OUT$ . Komponent  $IN$  obejmuje wierzchołki, z których istnieje ścieżka do największej silnej składowej spójności, ale brak ścieżki powrotnej. Komponent  $OUT$  obejmuje wierzchołki, do których prowadzi ścieżka z największej silnej składowej, jednak brak możliwości powrotu do niej.

Parametr	$WCC$	$SCC$	$IN$	$OUT$
Liczba	1	1	0	0
Największa składowa	3500	3500	–	–

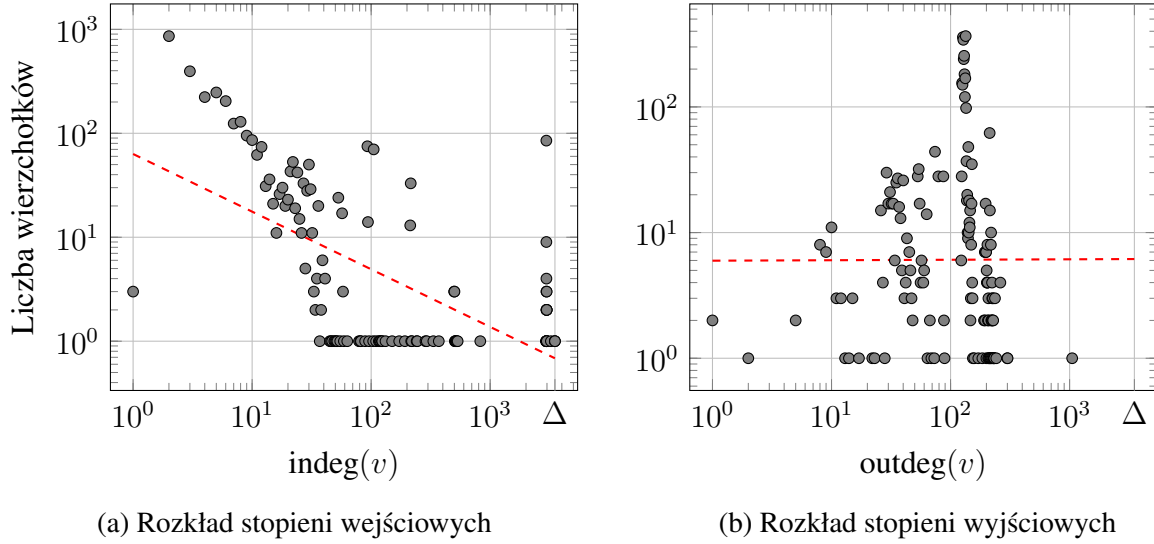
Tabela 2.2: Parametry dla składowych spójności.

Na podstawie danych zawartych w tabeli 2.2 wynika, że liczba silnych składowych spójności oraz słabych składowych spójności w grafie WWW wynosi 1, których liczba wierzchołków w największych składowych wynosi po 3500. Wynika z tego, że graf WWW jest silnie spójny, co oznacza istnienie ścieżek w obu kierunkach pomiędzy dowolnymi parami wierzchołków. Graf pozostaje spójny również po zignorowaniu kierunkowości krawędzi. W wyniku tego komponenty typu  $IN$  oraz  $OUT$  nie występują oddzielnie, a cały graf tworzy jednolitą strukturę. Dla domeny WWW graf  $SCC$ , w którym każda składowa to wierzchołek, składa się tylko z jednego elementu.

## 2.3 Rozkłady stopni

Stopień wierzchołka  $v$  grafu  $G$  oznaczany symbolem  $\deg(v)$  jest liczbą krawędzi incydenentnych z  $v$ . Maksymalny stopień grafu  $\Delta$  oznacza największy stopień spośród wszystkich wierzchołków. W grafie skierowanym rozróżnia się stopnie na wejściowe i wyjściowe. Stopień wejściowy  $\text{indeg}(v)$  oznacza liczbę krawędzi wchodzących do wierzchołka  $v$ . Stopień wyjściowy  $\text{outdeg}(v)$  oznacza liczbę krawędzi wychodzących z wierzchołka  $v$ .

Na rysunkach 2.3a i 2.3b przedstawiono rozkłady  $\text{indeg}(v)$  i  $\text{outdeg}(v)$  w zależności od liczby wierzchołków dla grafu WWW. Analizując rozkład  $\text{indeg}(v)$  można zauważyć wyraźny trend malejący. Wyjątek stanowi kilka wierzchołków o stopniach zbliżonych bądź równych  $\Delta$ , które pełnią rolę łączników z pozostałymi. Najwięcej stron jest dla  $\text{outdeg}(v)$  wynoszącego ok. 100, takie strony pełnią rolę hubów dla domeny.



Rysunek 2.3: Rozkłady stopni grafu WWW.

## 2.4 Ścieżki

Odległość  $d$  w grafie  $G$  pomiędzy dwoma wierzchołkami  $u$  i  $v$  jest definiowana jako długość najkrótszej ścieżki łączącej te dwa wierzchołki. Najkrótsza ścieżka to taka, która zawiera najmniejszą liczbę krawędzi. Średnia odległość między parami wierzchołków oblicza się poprzez zsumowanie długości wszystkich najkrótszych ścieżek między parami wierzchołków, a następnie podzielenie tej sumy przez liczbę takich par:

$$d_{\text{AVG}} = \frac{1}{n(n-1)} \sum_{u,v \in V} d(u, v)$$

Średnica grafu to największa odległość pomiędzy dwiema parami wierzchołków:

$$\text{Średnica} = \max_{u,v \in V} d(u, v)$$

Ekscentryczność wierzchołka  $v$  to największa odległość od  $v$  do dowolnego innego wierzchołka w grafie:

$$e(v) = \max_{u \in V} d(v, u)$$

Promień grafu to najmniejsza ekscentryczność wśród wszystkich wierzchołków w grafie:

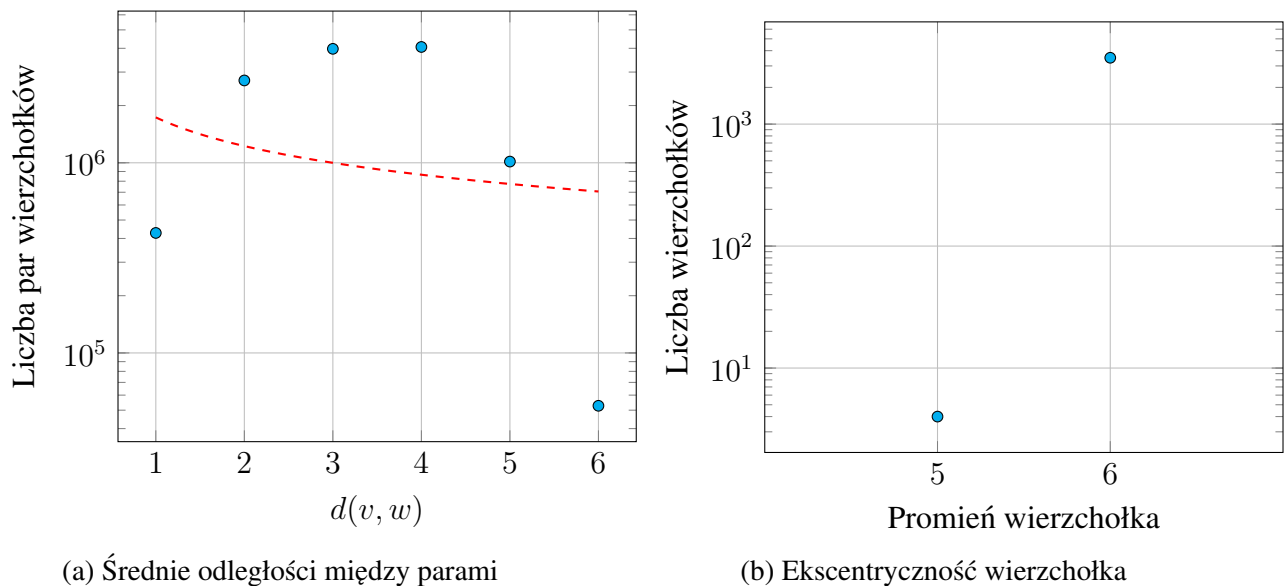
$$\text{Promień} = \min_v \left( \max_{u \in V} d(v, u) \right)$$

W tabeli 2.3 pokazano wyniki dla grafu WWW. Analizując średnią odległość można zauważyć, że przeciętnie dwie strony są od siebie oddalone o nieco ponad trzy skoki. Oznacza to, że dojście do dowolnej strony jest możliwe w niewielkiej liczbie przejść. Średnica wynosząca 6 wskazuje, że najbardziej oddalone od siebie strony można połączyć ścieżką o długości sześciu przejść. Z Rys. 2.4b wynika, że większość stron ma promień równy średnicy, co sugeruje względnie równomierne rozmieszczenie stron w strukturze domeny.

Parametr	Wartość
Średnia odległość	3,219
Średnica	6
Promień	5

Tabela 2.3: Parametry dotyczące odległości w badanym grafie.

Na Rys. 2.4a pokazano średnie odległości między parami wierzchołków. Największy udział mają pary oddalone o 3 i 4 jednostki, co świadczy o tym, że graf ma zwartą strukturę i jest zgodny ze średnią odległością w grafie.



Rysunek 2.4: Parametry odległości w grafie WWW.

## 2.5 Współczynniki klasteryzacji

Współczynniki klasteryzacji opisują, w jakim stopniu wierzchołki w grafie tworzą lokalne skupiska. Mierzą tendencję sąsiednich wierzchołków do bycia wzajemnie połączonymi.

Dla wierzchołka  $i$  lokalny współczynnik klasteryzacji  $C_i$  definiowany jest w następujący sposób:

$$C_i = \frac{(\text{liczba par sąsiadów wierzchołka } i, \text{ które są połączone})}{(\text{liczba wszystkich par sąsiadów wierzchołka } i)}$$



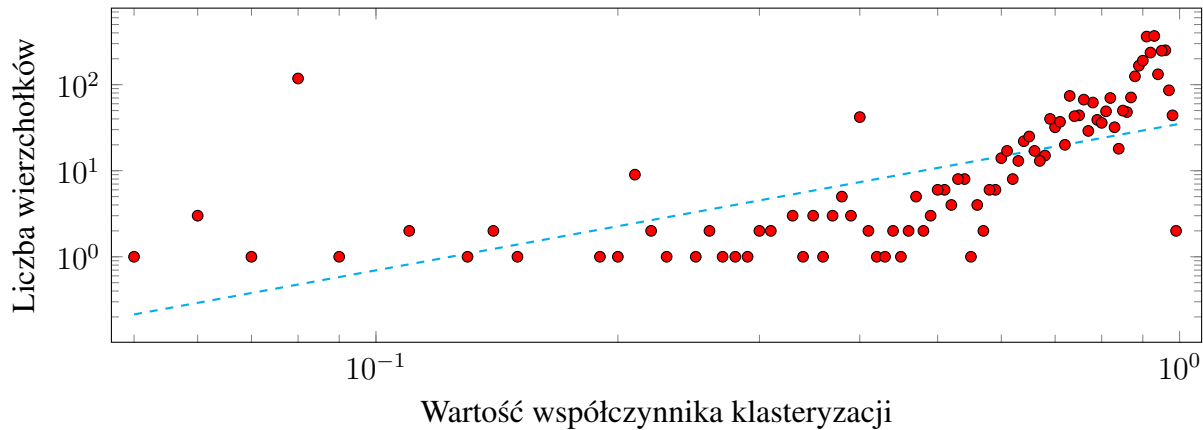
Globalny współczynnik klasteryzacji mierzy ogólną tendencję całego grafu do tworzenia trójkątów i jest definiowany jako stosunek trzykrotności liczby trójkątów w grafie do liczby połączonych trójek:

$$C = \frac{3 \times (\text{liczba trójkątów})}{(\text{liczba połączonych trójek})}$$

Dla badanego grafu globalny współczynnik klasteryzacji wyniósł:

$$C = 0,133$$

Na rysunku 2.5 przedstawiono histogram wartości lokalnego współczynnika klasteryzacji dla grafu WWW. Analiza wykresu i globalnej wartości pokazuje, że graf jest luźno połączony globalnie, ale lokalnie zawiera zwarte grupy.



Rysunek 2.5: Rozkład wartości współczynnika klasteryzacji.

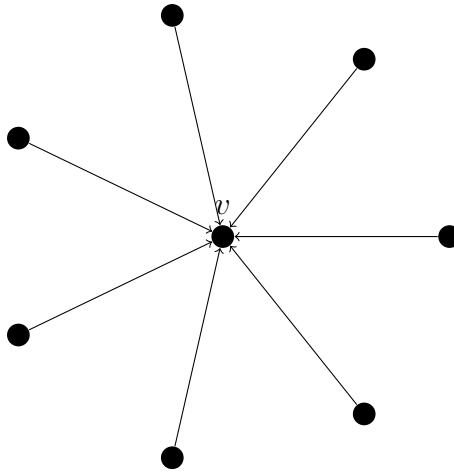
## 2.6 Spójność wierzchołkowa

Niech  $G = (V, E)$  będzie grafem, gdzie  $V(G)$  to zbiór wierzchołków, a  $E(G)$  to zbiór krawędzi. Spójność wierzchołkowa  $\kappa(G)$  grafu  $G$  to minimalna liczba wierzchołków, które trzeba usunąć z grafu aby stał się niespójny, czyli nie będzie istnieć ścieżka między wszystkimi parami wierzchołków w grafie po ich usunięciu.

Dla badanego grafu WWW spójność wierzchołkowa wynosi:

$$\kappa(G) = 1$$

Oznacza to, że istnieją jedna strona, która spaja dużą część domeny. Tym elementem spajającym w domenie jest strona główna. Jej niedostępność, np. spowodowana przez awarię lub atak, sprawi, że duża część serwisu stanie się niedostępna albo mocno rozczłonkowana. Badana domena ma podobny charakter jak graf pokazany na Rys. 2.6, gdzie usunięcie wierzchołka  $v$  spowoduje jego rozspójnienie.



Rysunek 2.6: Przykład grafu gdzie  $\kappa = 1$ .

## 2.7 Pagerank

Algorytm PageRank przypisuje każdej ze stron internetowych pewien współczynnik, który określa wagę tej strony względem innych stron. Współczynnik ten jest związany z prawdopodobieństwem, że internauta klikając w sposób losowy w odnośnik, ostatecznie trafi na rozpatrywaną stronę. Jeśli strona  $v$  ma wiele wchodzących odnośników, to jego PageRank rośnie proporcjonalnie do sumy wartości PageRank stron linkujących do  $v$ , podzielonej przez liczbę ich wychodzących linków. Aby zapewnić zbieżność i umożliwić przejścia w przypadkach stron bez wyjścia, wprowadzany jest współczynnik tłumienia  $\alpha$ . Ostatecznie, w każdej iteracji, nowy PageRank dla strony  $v$  oblicza się według wzoru:

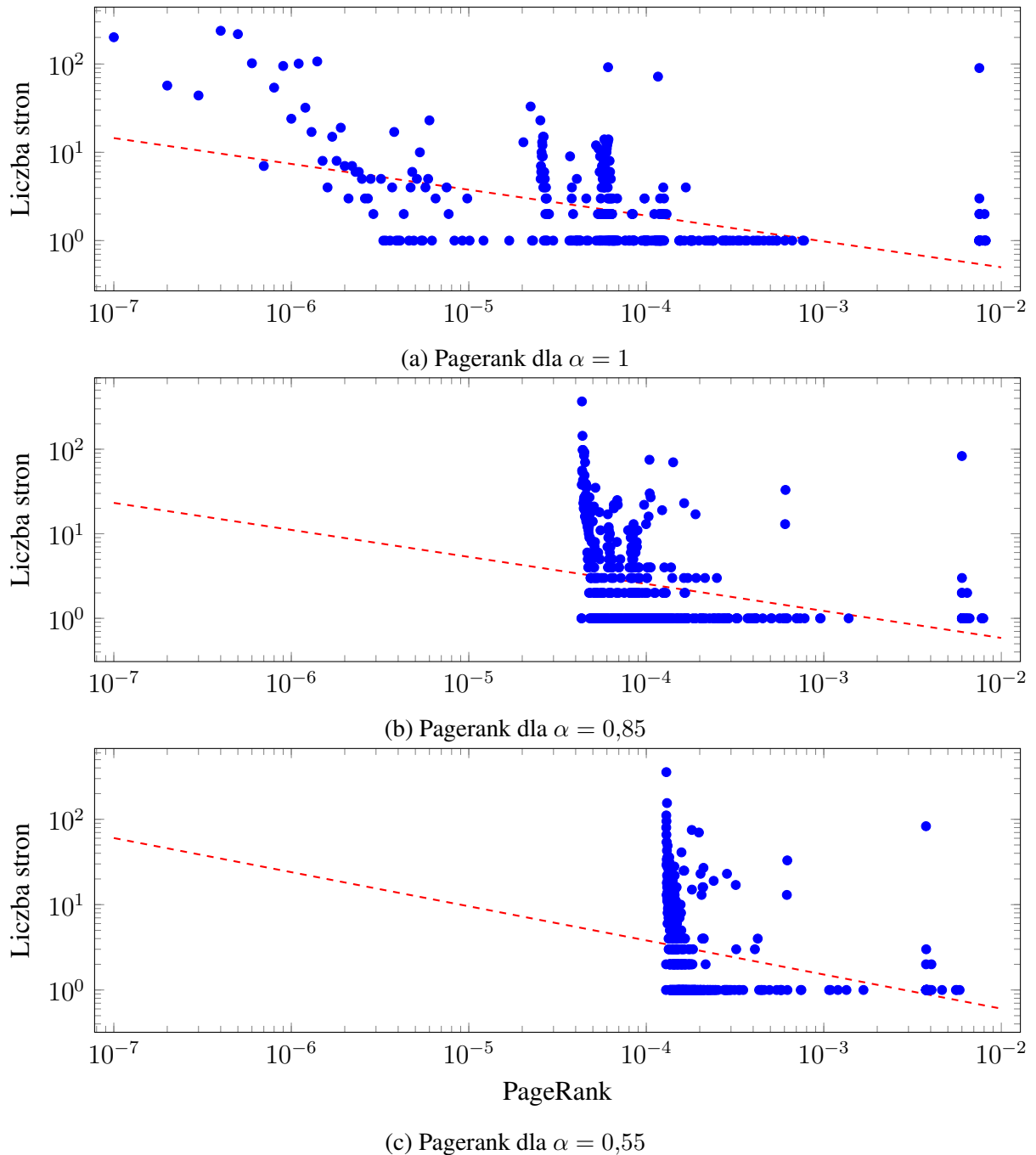
$$\mathcal{PR}(v) = \frac{1 - \alpha}{N} + \alpha \sum_{u \in \text{IN}(v)} \frac{\mathcal{PR}(u)}{\text{outdeg}(u)}$$

W tabeli 2.4 przedstawiono liczbę iteracji potrzebnych do osiągnięcia zbieżności implementowanego algorytmu PageRank przy różnych wartościach współczynnika tłumienia  $\alpha$ . Zbieżność została określona jako osiągnięcie tolerancji błędu  $10^{-6}$ . Obniżenie wartości  $\alpha$  skutkuje szybszą stabilizacją wartości PageRank i zmniejszeniem liczby iteracji.

Współczynnik $\alpha$	Liczba iteracji
1	84
0,85	35
0,55	14

Tabela 2.4: Liczba iteracji algorytmu dla różnych  $\alpha$ .

Na Rys. 2.7 przedstawiono rozkład wartości PageRank w zależności od liczby stron dla różnych wartości parametru  $\alpha$ . W przypadku ustawienia parametru tłumienia  $\alpha = 1$ , algorytm PageRank nie wykazuje zbieżności ze względu na brak mechanizmu losowego przejścia, co widać na Rys. 2.7a. Dopiero wprowadzenie tłumienia zapewnia globalną zbieżność poprzez umożliwienie losowych przeskoków między wszystkimi wierzchołkami grafu. Dla  $\alpha = 0,85$  i  $\alpha = 0,55$  PageRank osiąga zbieżność, przy czym  $\alpha = 0,55$  wykres przedstawia jeszcze większe zagęszczenie przy wartości  $\mathcal{PR} = 10^{-4}$ . Wartość  $\alpha = 0,85$  wydaje się być dobrym równoważnikiem pomiędzy podążaniem za odnośnikami a losowością.



Rysunek 2.7: Rozkład wartości PageRank.

## 3. Odporność na ataki

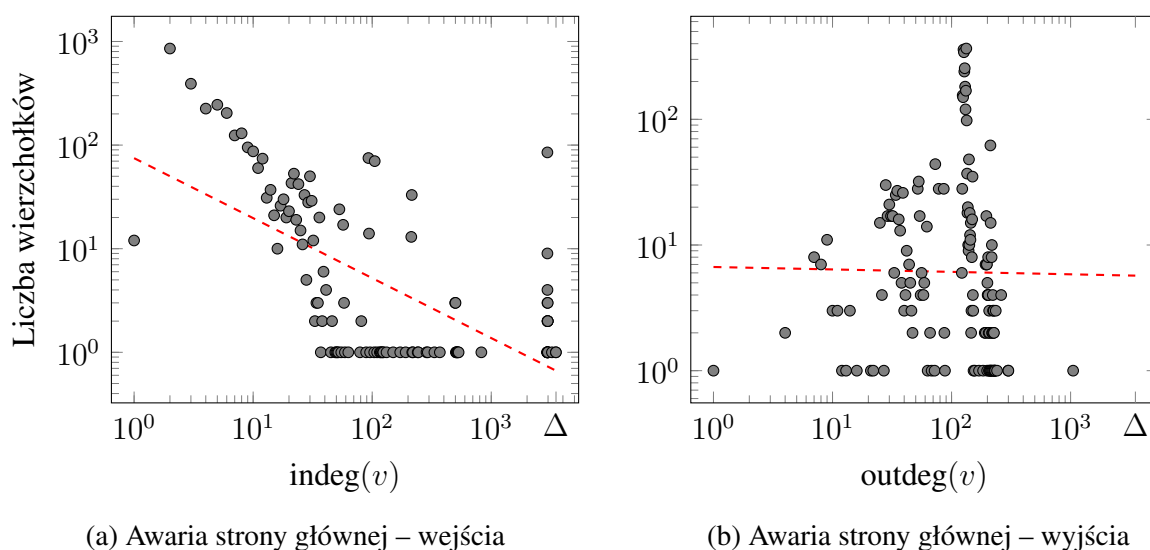
W ramach analizy odporności sieci na awarie i ataki, przeprowadzono symulacje, które obejmują zarówno losowe usuwanie wierzchołków, jak i usuwanie wierzchołków o maksymalnym stopniu. Celem eksperymentu było zbadanie, jak różne poziomy usunięć wierzchołków wpływają na strukturę sieci oraz jej parametry, takie jak składowe spójności, rozkład stopni wierzchołków i odległości między wierzchołkami.

### 3.1 Awaria stron

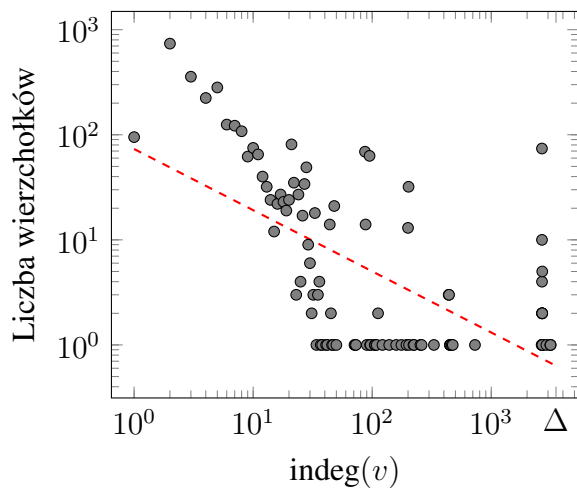
W pierwszej części eksperymentu przeprowadzono awarię na sieci, usuwając wierzchołki w sposób losowy. W tym przypadku symulowano usunięcie strony głównej, 10%, 20% i 50% losowych wierzchołków z sieci. Dzięki temu można było zaobserwować, jak zanik wierzchołków wpływa na integralność sieci, zmieniając licznosc składowych spójności oraz rozkład stopni.

#### 3.1.1 Rozkłady stopni

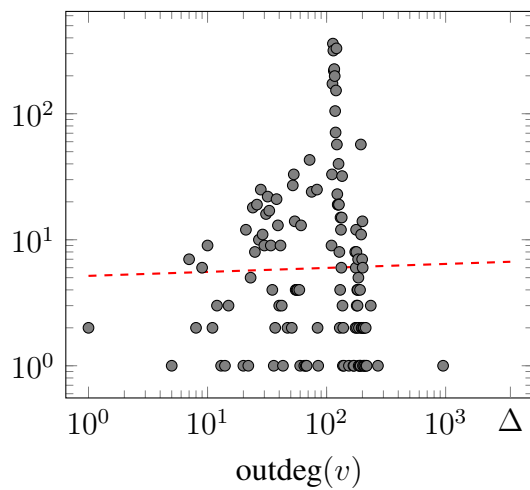
Na Rys. 3.1, 3.2, 3.3 i 3.4 przedstawiono rozkłady stopni dla różnych scenariuszy awarii.



Rysunek 3.1: Rozkłady wejść i wyjść po awarii strony głównej.

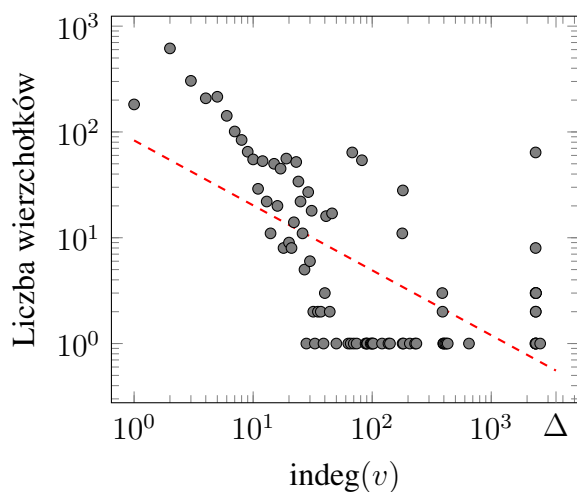


(a) Awaria 10% – wejścia

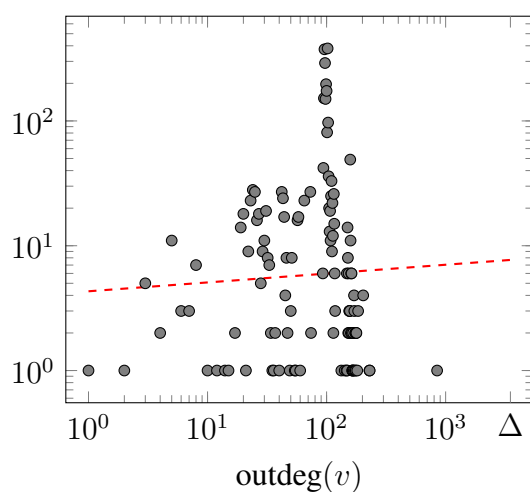


(b) Awaria 10% – wyjścia

Rysunek 3.2: Rozkłady wejść i wyjść po awarii 10% stron.

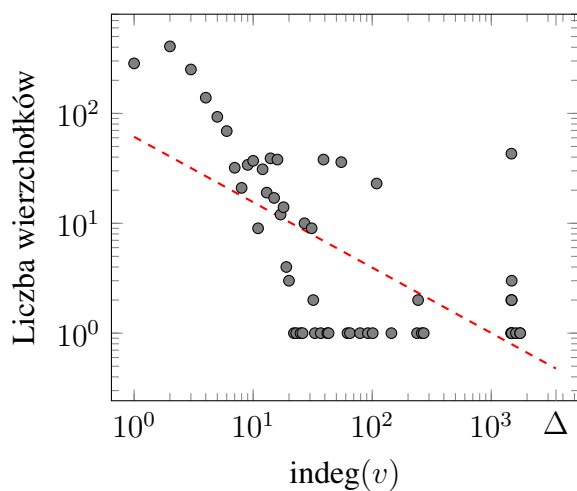


(a) Awaria 20% – wejścia

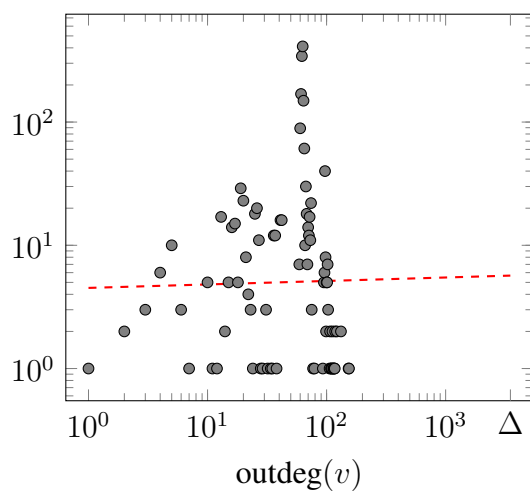


(b) Awaria 20% – wyjścia

Rysunek 3.3: Rozkłady wejść i wyjść po awarii 20% stron.



(a) Awaria 50% – wejścia



(b) Awaria 50% – wyjścia

Rysunek 3.4: Rozkłady wejść i wyjść po awarii 50% stron.

### 3.1.2 Składowe spójności

Tabela 3.1 pokazuje, jak różne poziomy awarii wpływają na spójność grafu WWW. Początkowo graf tworzy jedną silnie i słabo spójną składową, bez komponentów  $IN$  ani  $OUT$ . Już awaria strony głównej powoduje rozpad  $SCC$  na 19 części, choć  $WCC$  pozostaje niemal nienaruszona. Wraz ze wzrostem poziomu awarii, liczba  $SCC$  gwałtownie rośnie, a rozmiar największej  $SCC$  wyraźnie maleje. Komponent  $IN$  znacznie się powiększa, wskazując na węzły prowadzące do  $SCC$ , ale nieosiągalne z niej. Komponent  $OUT$  pojawia się jedynie lokalnie i zanika, co sugeruje utratę kierunkowej łączności wychodzącej.

Parametr	Domena WWW	Awaria strony głównej	Awaria 10%	Awaria 20%	Awaria 50%
Liczba $SCC$	1	19	159	330	717
Liczba $WCC$	1	1	1	1	1
Największa składowa $SCC$	3500	3480	2863	2323	756
Największa składowa $WCC$	3500	3499	3150	2800	1750
Liczba $IN$	0	16	287	476	994
Liczba $OUT$	0	3	0	1	0

Tabela 3.1: Podsumowanie składowych spójności dla awarii.

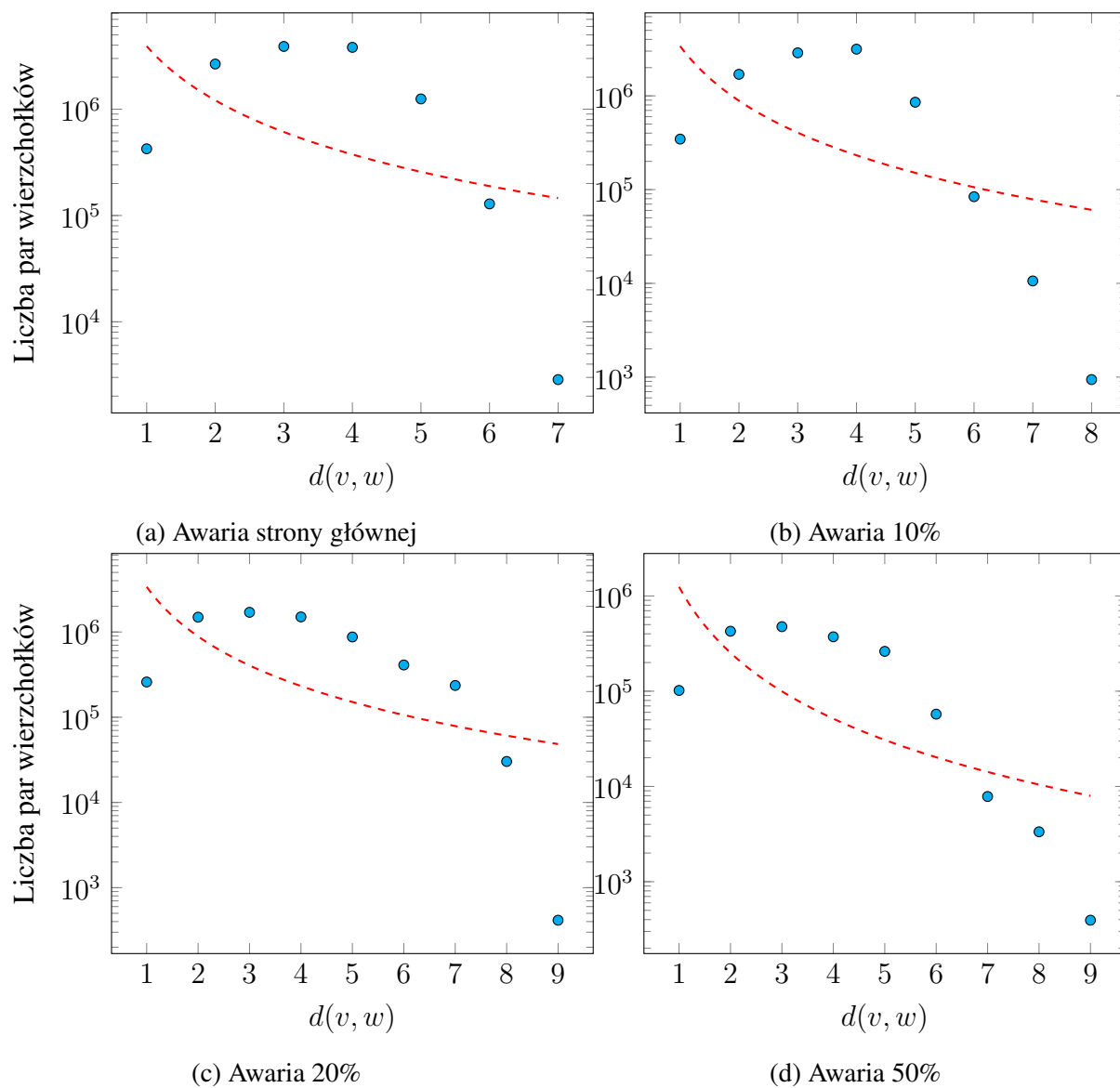
### 3.1.3 Odległości

W tabeli 3.2 przedstawiono podsumowanie scenariuszy awarii dla odległości. Po awarii strony głównej średnia odległość wzrasta nieznacznie do 3,26. Przy awarii 10% losowych stron średnia odległość wzrasta jeszcze bardziej do 3,31. Awaria 20% losowych stron prowadzi do dalszego wzrostu średniej odległości do 3,55. Przy awarii 50% wierzchołków średnia odległość wynosi 3,27, co jest niższe niż przy awarii 20%. Wynika to z faktu, że graf po usunięciu połowy wierzchołków wciąż ma mieć fragmenty, które są wciąż względnie blisko siebie.

Przypadek	$V(G)$	$E(G)$	$d_{AVG}$	$\kappa(G)$
Domena WWW	3500	431129	3.22	1
Awaria strony głównej	3499	427480	3.26	0
Awaria 10%	3150	348947	3.31	0
Awaria 20%	1750	103587	3.55	0
Awaria 50%	3150	348947	3.27	0

Tabela 3.2: Podsumowanie odległości dla awarii.

Na Rys. 3.5 przedstawiono średnie odległości między parami wierzchołków dla scenariuszy awarii.



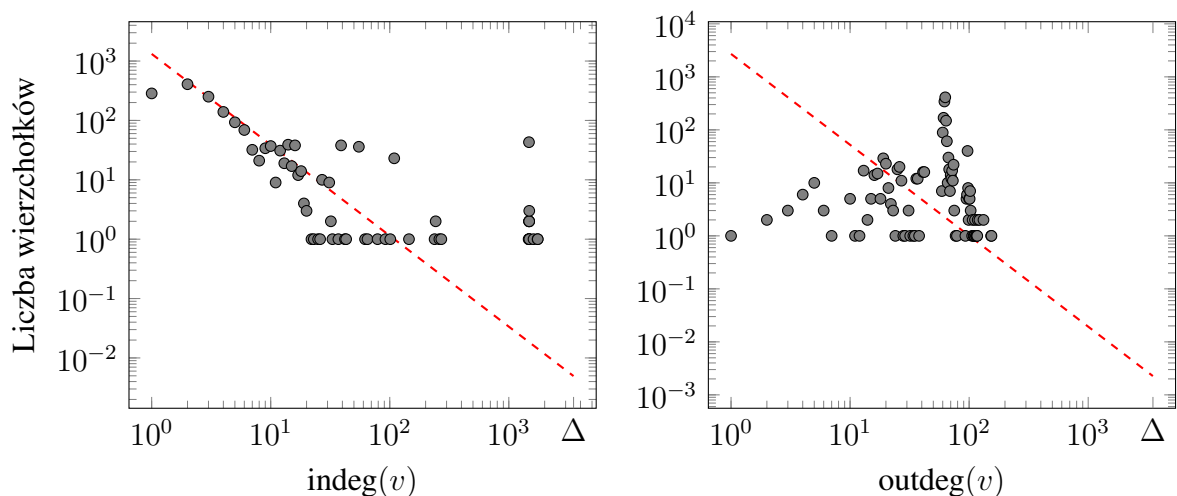
Rysunek 3.5: Średnie odległości między parami wierzchołków.

## 3.2 Ataki

W drugiej części eksperymentu przeprowadzono atak na sieć, polegający na usunięciu wierzchołków o maksymalnym stopniu. Usunięto 10%, 20% i 50% wierzchołków, aby ocenić, jak bardziej ukierunkowane usuwanie wierzchołków (odpowiadających kluczowym węzłom w sieci) wpływa na jej strukturalną stabilność i spójność.

### 3.2.1 Rozkłady stopni

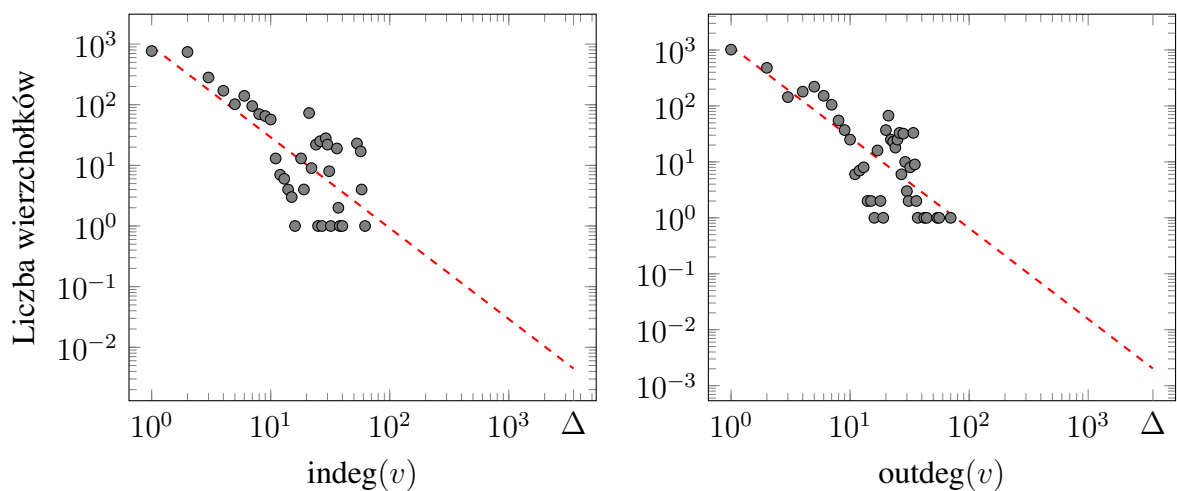
Na Rys. 3.6, 3.7 i 3.8 przedstawiono rozkłady stopni dla różnych scenariuszy awarii.



(a) Rozkład stopni wejściowych dla grafu

(b) Rozkład stopni wyjściowych dla grafu

Rysunek 3.6: Atak na 10% stron.

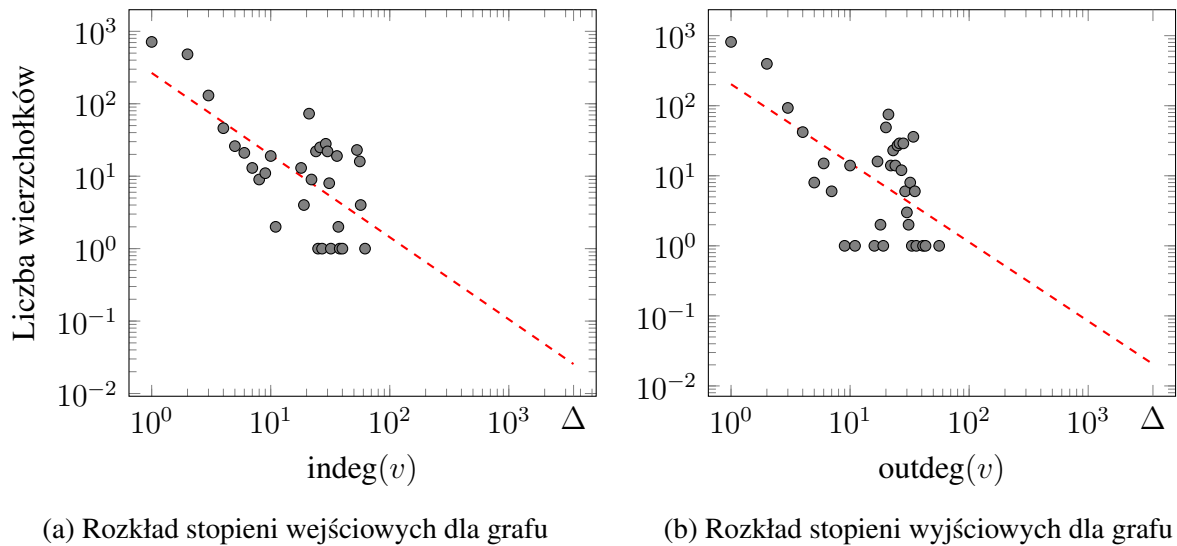


(a) Rozkład stopni wejściowych dla grafu

(b) Rozkład stopni wyjściowych dla grafu

Rysunek 3.7: Atak na 20% stron.





Rysunek 3.8: Atak na 50% stron.

### 3.2.2 Składowe spójności

Tabela 3.3 przedstawia wpływ ataków ukierunkowanych na strukturę spójności grafu WWW. Już przy 10% ataku liczba  $SCC$  rośnie drastycznie do 1290, a największa  $SCC$  maleje ponad dwukrotnie. Dalsze ataki powodują niemal całkowity rozpad silnej spójności — największa  $SCC$  ma tylko 61 wierzchołków przy 20% i 50% ataku.  $WCC$  również ulega fragmentacji, szczególnie przy 20% i 50%, co odróżnia atak od awarii. Komponenty  $IN$  i  $OUT$  pojawiają się już przy 10% ataku, ale szybko zanikają. Ataki silnie degradują zarówno kierunkową, jak i ogólną strukturę połączeń w sieci.

Parametr	Domena WWW	Atak 10%	Atak 20%	Atak 50%
Liczba $SCC$	1	1290	1857	1263
Liczba $WCC$	1	42	626	690
Największa składowa $SCC$	3500	1310	61	61
Największa składowa $WCC$	3500	3027	1794	436
Liczba $IN$	0	461	2	2
Liczba $OUT$	0	609	164	30

Tabela 3.3: Podsumowanie składowych spójności dla ataków.

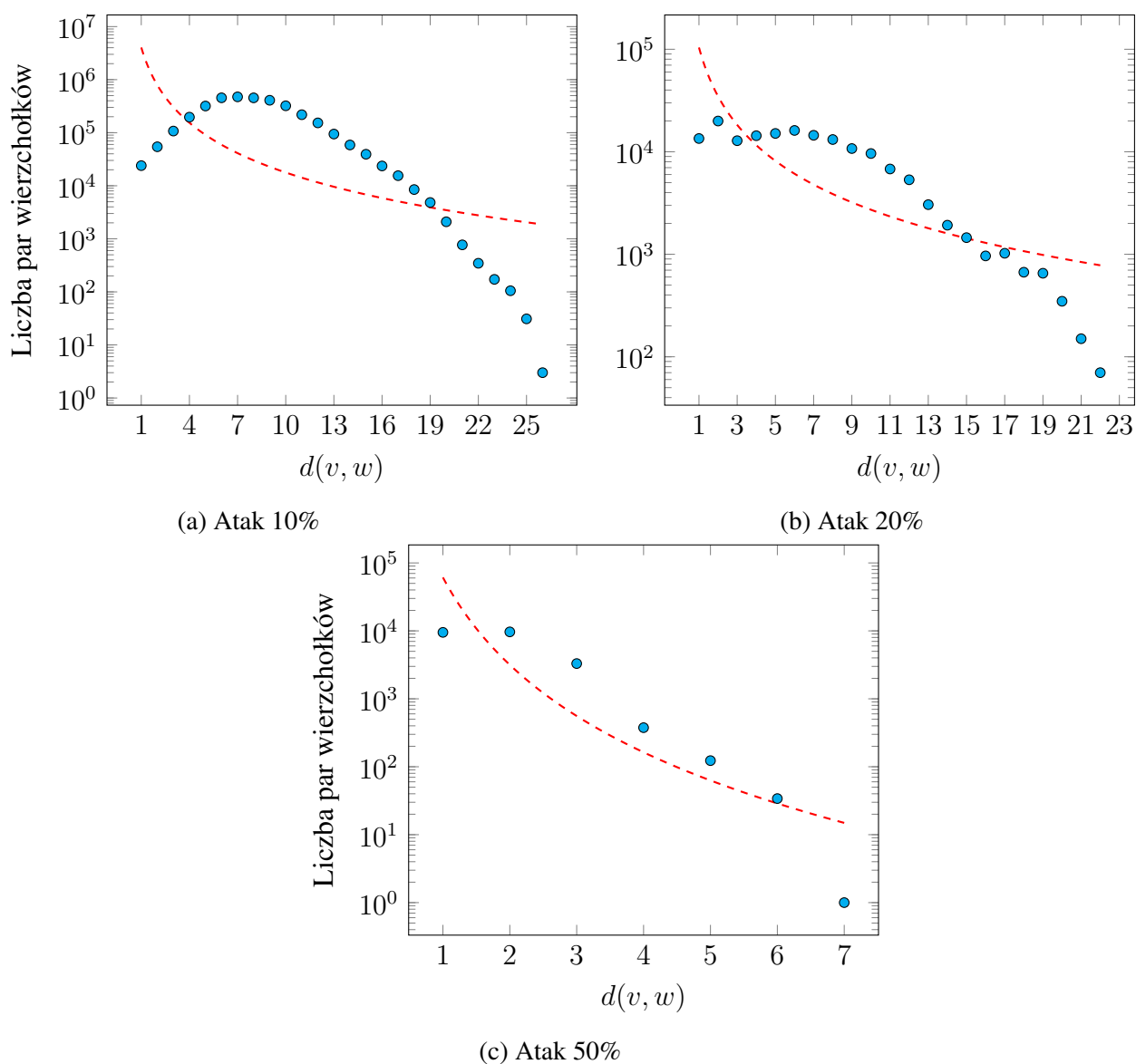
### 3.2.3 Odległości

W tabeli 3.4 przedstawiono podsumowanie scenariuszy ataków. Dla ataku na 10% losowych wierzchołków średnia odległość znacząco wzrasta do 7,94. To oznacza, że usunięcie tych wierzchołków miało poważny wpływ na strukturę grafu, niszcząc spójność i oddalając wierzchołki od siebie, co widać na Rys. 3.9a. Po ataku na 20% wierzchołków średnia odległość zmniejsza się do 6,27, zaś po ataku 50% spada znacząco do 1,78. Graf wciąż zachował jakąś formę lokalnej spójności, mimo że usunięcie połowy wierzchołków wywołało ogromne zmiany. W wyniku ataków graf został rozbity na mniejsze fragmenty, co potwierdza to skuteczność tych ataków.

Przypadek	$V(G)$	$E(G)$	$d_{AVG}$	$\kappa(G)$
Domena WWW	3500	431129	3.22	1
Atak 10%	3150	27087	7.94	0
Atak 20%	2800	16266	6.27	0
Atak 50%	1750	11246	1.78	0

Tabela 3.4: Podsumowanie odległości dla ataków.

Na Rys. 3.9 przedstawiono rozkłady odległości między parami wierzchołków dla awarii.



Rysunek 3.9: Średnie odległości między parami wierzchołków.

### 3.3 Wnioski

Graf WWW jest stosunkowo odporny na awarie, usunięcie strony głównej i 10% wierzchołków nie wpływa drastycznie na jego strukturę, ale jest bardzo wrażliwy na ataki, które prowadzą do szybkiej utraty spójności, znacznego zwiększenia średniej odległości między wierzchołkami i liczby składowych spójności. Spójność wierzchołkowa spada do 0 we wszystkich przypadkach awarii i ataków, co pokazuje spajający charakter strony głównej.

# Spis rysunków

1.	Wizualizacja grafu wybranej domeny dla 50 stron. . . . .	
1.1.	Przejście po grafie wg. algorytmu BFS. . . . .	4
1.2.	Czas wykonania programu w zależności od liczby wątków. . . . .	4
2.1.	Porównanie grafu prostego i digrafu. . . . .	5
2.2.	Podział na $WCC$ i $SCC$ . . . . .	6
2.3.	Rozkłady stopni grafu WWW. . . . .	7
2.4.	Parametry odległości w grafie WWW. . . . .	8
2.5.	Rozkład wartości współczynnika klasteryzacji. . . . .	9
2.6.	Przykład grafu gdzie $\kappa = 1$ . . . . .	10
2.7.	Rozkład wartości PageRank. . . . .	11
3.1.	Rozkłady wejść i wyjść po awarii strony głównej. . . . .	12
3.2.	Rozkłady wejść i wyjść po awarii 10% stron. . . . .	13
3.3.	Rozkłady wejść i wyjść po awarii 20% stron. . . . .	13
3.4.	Rozkłady wejść i wyjść po awarii 50% stron. . . . .	13
3.5.	Średnie odległości między parami wierzchołków. . . . .	15
3.6.	Atak na 10% stron. . . . .	16
3.7.	Atak na 20% stron. . . . .	16
3.8.	Atak na 50% stron. . . . .	17
3.9.	Średnie odległości między parami wierzchołków. . . . .	19

# Spis tabel

2.1. Podstawowe parametry badanego grafu. . . . .	5
2.2. Parametry dla składowych spójności. . . . .	6
2.3. Parametry dotyczące odległości w badanym grafie. . . . .	8
2.4. Liczba iteracji algorytmu dla różnych $\alpha$ . . . . .	10
3.1. Podsumowanie składowych spójności dla awarii. . . . .	14
3.2. Podsumowanie odległości dla awarii. . . . .	14
3.3. Podsumowanie składowych spójności dla ataków. . . . .	17
3.4. Podsumowanie odległości dla ataków. . . . .	18