Generator sieci małego świata

Iryna Mikutskaya

W ramach zadania „Generator sieci małego świata" powinien być zaimplementowany model Wattsa-Strogatza. Program powinien zapisywać do plików tekstowych:

- wygenerowany graf do pliku .csv,
- rozkład stopni wierzchołków,
- rozkład współczynnika grupowania.

Za pomocą generatora musi być sprawdzone, czy ww. rozkłady zgadzają się z przewidywaniami teoretycznymi.

Program powinien być przenośny.

Do napisania programu będzie użyte środowisko MATLAB.

Parametrami początkowymi w generatorze sieci małego świata będą:

n - liczba wierzchołków,

d - stopień wierzchołków (liczba parzysta),

p - prawdopodobieństwo przełączenia krawędzi.

Graf będzie przedstawiony w postaci macierzy sąsiedztwa *nxn*. Na początku w macierzy będą same zera, potem będą dodawane połączenia między krawędziami, wpisując do macierzy jedynki w taki sposób, że *d/2* jedynek będzie dodane bezpośrednio po rozpatrywanym wierzchołku i *d/2* jedynek przed rozpatrywanym wierzchołkiem.

Pokażemy jak to się robi na przykładzie.

Mamy sieć z $n=10$, $d=4$.

Tworzymy macierz 10x10 w której są same zera.

Potem tworzymy na jej podstawie macierz sąsiedztwa wpisując do niej 1 jak opisane jest wyżej.

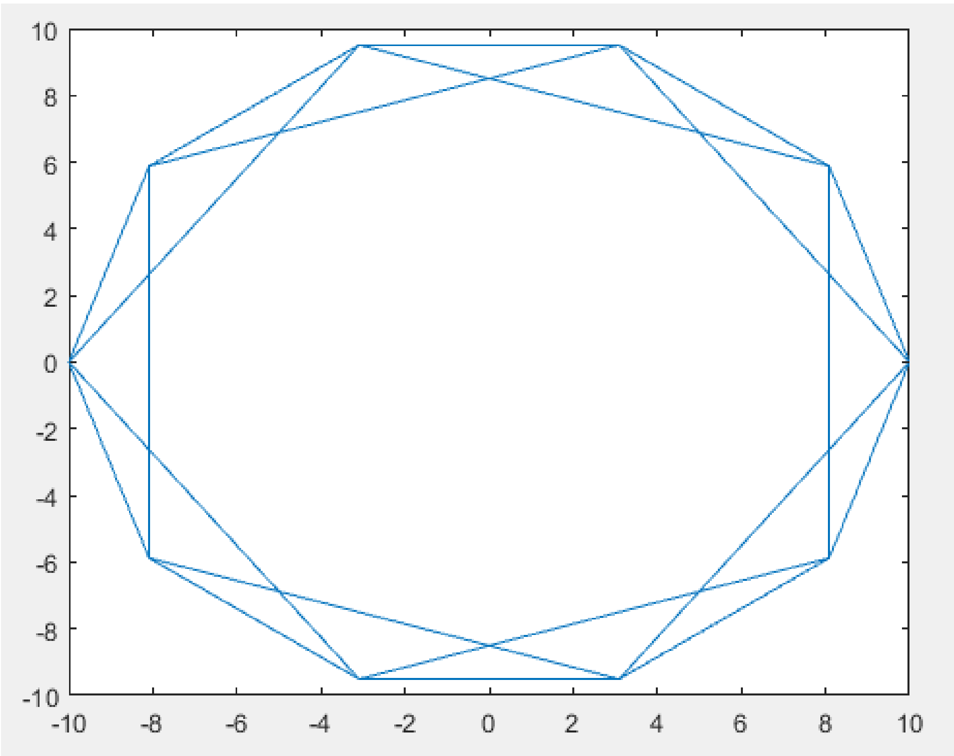
tab. 1. Macierz sąsiedztwa

	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	0	1	1	0	0	0	0	0	1	1
2	1	0	1	1	0	0	0	0	0	1
3	1	1	0	1	1	0	0	0	0	0
4	0	1	1	0	1	1	0	0	0	0
5	0	0	1	1	0	1	1	0	0	0
6	0	0	0	1	1	0	1	1	0	0
7	0	0	0	0	1	1	0	1	1	0
8	0	0	0	0	0	1	1	0	1	1
9	1	0	0	0	0	0	1	1	0	1
10	1	1	0	0	0	0	0	1	1	0

Kolorem żółtym są oznaczone w tabeli wierzchołki, które będziemy rozpatrywać. W te kratki jedynki nie dodajemy, bo sieć nie ma pętli własnych. *4/2(d/2)* jedynek dodaliśmy po i przed żółtymi kratkami (tab. 1).

Przy takim rozkładzie macierz jest symetryczna.

Wynikiem będzie graf, pokazany na rys. 1.



rys. 1. Graf z parametrami n=10, d=4.

W macierzy początkowej są przełączane krawędzie w następujący sposób. Dla każdego węzła jest brana krawędź (n_i; n_j) z indeksami $i<j$ i przełączana z prawdopodobieństwem *p* tak, że w wyniku otrzymamy krawędź (n_i; n_k). Węzeł, od którego będzie odpięta krawędź, reprezentowany jest jedynką w wierszu macierzy sąsiedztwa w miejscu z indeksem o 1 większym od indeksu rozpatrywanego węzła. Węzeł do którego zostanie podpięta krawędź, jest wybierany losowo.

W macierzy wygląda to tak: każdy wiersz macierzy jest analizowany oddzielnie. Brana jest pierwsza idącą po rozpatrywanym wierzchołku jedynka w wierszu (pierwsza jedynka po żółtej kratce w tab.1), z prawdopodobieństwem *p* jest ona przeniesiona w jedno z miejsc z zerami w tym samym wierszu macierzy. A na poprzednim jej miejscu będzie ustawione 0. Następnie jest analizowany drugi wiersz macierzy i tak do ostatniego wiersza.

Po przejściu tak przez wszystkie wierzchołki, powtarzane to jest od nowa, tylko tym razem brana jest jedynka o indeksie o jeden większym, niż w poprzedniej iteracji (brany jest nie najbliższy sąsiad, a o 1 dalszy, tzn. następny najbliższy sąsiad). To działanie jest kontynuowane dopóki każda oryginalna krawędź nie zostanie rozpatrzona jeden raz. Ponieważ w całej sieci są *nd/2* krawędzie, proces przewijania zatrzymuje się po około *d/2* okrążeniach.[1]

Jak już mamy końcową macierz, to można na jej podstawie otrzymać parametry takie jak rozkład stopni wierzchołków i współczynnik grupowania.

Dla tego, żeby dowiedzieć się jakie stopnie mają wierzchołki sieci, będą sumowane oddzielnie jedynki albo w każdym wierszu, albo w każdej kolumnie. A potem będzie liczona ilość wierzchołków otrzymanych stopni.

Współczynnik grupowania będzie obliczony dla każdego wierzchołka oddzielnie. Współczynnik grupowania wierzchołka to stosunek liczby E_i istniejących krawędzi między sąsiadami wierzchołka do wszystkich możliwych krawędzi między tymi sąsiadami $k_i(k_i - 1)/2$. Jest on zatem równy

$$C_i=2E_i/k_i(k_i-1), \tag{1}$$

gdzie $0 \leq E_i \leq k_i(k_i - 1) / 2$, dzięki czemu $0 \leq C_i \leq 1$. Jak widzimy, współczynnik gronowania opisuje prawdopodobieństwo, że pierwsi sąsiedzi węzła i są względem siebie również pierwszymi sąsiadami [2].

Po obliczeniu współczynnika grupowania dla każdego wierzchołka, można znaleźć taki współczynnik dla całej sieci, to znaczy wartość C_i uśrednioną po wszystkich węzłach.

$$C = \langle C_i \rangle. \tag{2}$$

Wynik zapiszemy do pliku tekstowego.

Za pomocą generatora będzie sprawdzane, czy ww. rozkłady zgadzają się z przewidywaniami teoretycznymi. Wynikiem porównania będą wykresy.

Implementacja modelu Wattsa-Strogatza

Na początku algorytmu są podane wartości zmiennych *n, d, p* – liczba wierzchołków, stopień wierzchołków i prawdopodobieństwo przełączenia krawędzi. Następnie, opierając się na ww wartości, jest budowana symetryczna macierz sąsiedztwa.

Przełączenie krawędzi w macierzy jest opisane wyżej. W algorytmie to jest zaimplementowane w następujący sposób. Każda krawędź w sieci powinna być przeanalizowana tylko jeden raz. Na pierwszym okrążeniu analizujemy pierwszych najbliższych sąsiadów, na drugim – drugich i tak dalej. Takich okrążeń powinno być *d/2,* algorytm analizowania krawędzi będzie znajdował się w pętli *for b=1:d/2*. W każdym okrążeniu przechodzimy przez wszystkie wierzchołki (*n*). W zmienną *ind1* jest zapisywany indeks *d*-go sąsiada. Następnie jest generowana wartość losowa, która jest porównywana z wartością zmiennej *p* (prawdopodobieństwo przełączenia krawędzi). Jeśli wygenerowana wartość jest większa od *p*, to krawędź nie będzie przełączona, przechodzimy do następnego wierzchołka i analizujemy jego *d*-go sąsiada. Jeśli wygenerowana wartość jest mniejsza albo równa *p*, to wypełniamy szereg działań. Jest tworzona tablica z indeksami wierzchołków, do których może być przełączona krawędź. Losowo jest wybierany z tej tablicy wierzchołek, do którego krawędź będzie podpięta (nie może tworzyć pętli własnej i krawędzi współbieżnych w sieci). W macierzy sąsiedztwa ustawiane jest zero w miejscu dla węzła, od którego krawędź była odłączona, i ustawiamy jedynkę w miejsce dla wierzchołka, do którego ta krawędź była podpięta. I tak jest zrobione *d/2* okrążenia.

Po przeanalizowaniu całej sieci, jest liczone ile wierzchołków każdego stopnia ma graf (badamy rozkład stopni grafu). Na początku zapisujemy do tablicy wszystkie stopnie wierzchołków (stopień jest równy ilości jedynek w poszczególnym wierszu). Jest znajdowany minimalny i maksymalny stopień i w pętli jest liczona ilość wierzchołków każdego stopnia. Wszystkie pary *stopień–ilość wierzchołków* są zapisywane do pliku tekstowego *rstopni.txt* w postaci tablicy.

Następnie jest wyliczany współczynnik grupowania dla poszczególnych wierzchołków zgodnie ze wzorem (1), i dla całej sieci (wzór (2)). W programie to wygląda następująco:

```
wspgr=zeros(1,n);
for i=1:n
  ci=0;
  e=0;
  for j=1:n
    if M(i,j)==1
      j2=j+1;
      for j1=j2:n
        if M(i,j1)==1 && M(j,j1)==1
          e=e+1;
        end
      end
    end
  end
```

```
  ci=(2*e)/(d*(d-1));
  wspgr(i)=ci;
end
ws=0;
for i=1:n
ws=ws+wspgr(i);
end
wsgrs=ws/n;
```

Do obliczenia tego współczynnika dla każdego z wierzchołków są sumowane jedynki w wierszu (k_i) i dla każdej z tych jedynek jest sprawdzane, czy są połączone między sobą odpowiednie węzły. Np. gdy jedynki są w pierwszym wierszu macierzy pod indeksami {1,3}, {1,7} i {1,9}, to jest sprawdzane czy pod indeksami {3,7}, {3,9} i {7,9} też są 1. Jeśli są, to jest zapamiętana ich ilość w zmiennej (E_i) - liczba istniejących krawędzi między sąsiadami wierzchołka.

Współczynnik grupowania całej sieci to wartość C_i uśredniona po wszystkich węzłach.

Wyniki są zapisane do pliku *wspgrup.txt*.

Macierz sąsiedztwa wygenerowanego grafu jest zapisana do pliku *M.csv*.

Dla sprawdzenia poprawności działania zaimplementowanego generatora będziemy sprawdzać otrzymane wyniki dla rozkładu stopni oraz zależność współczynnika grupowania od prawdopodobieństwa przełączania krawędzi z teoretycznie przewidywanymi wartościami.

Przewidywania teoretyczne rozkładu stopni wierzchołków są obliczane według wzoru 3.[4]

$$P(d=k) \approx \frac{e^{-\bar{d}}\bar{d}^k}{k!}. \tag{3}$$

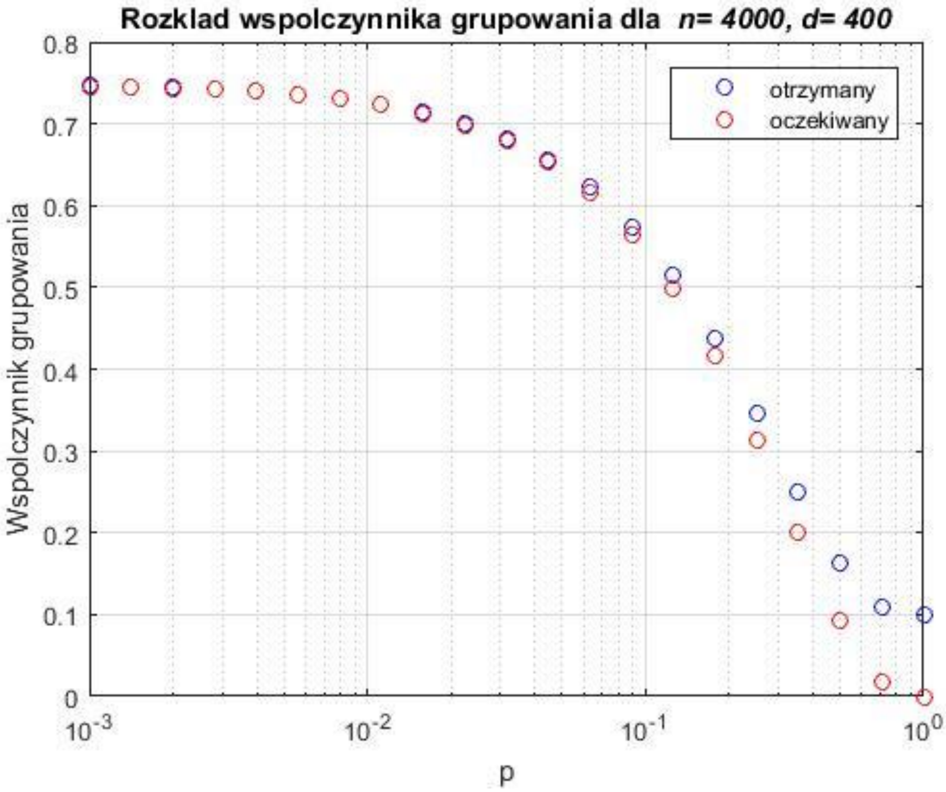
Przewidywania teoretyczne współczynnika grupowania są obliczane według wzoru 4[3].

$$\tilde{C}(p) = \frac{3(k-1)}{2(2k-1)}(1-p)^3. \tag{4}$$

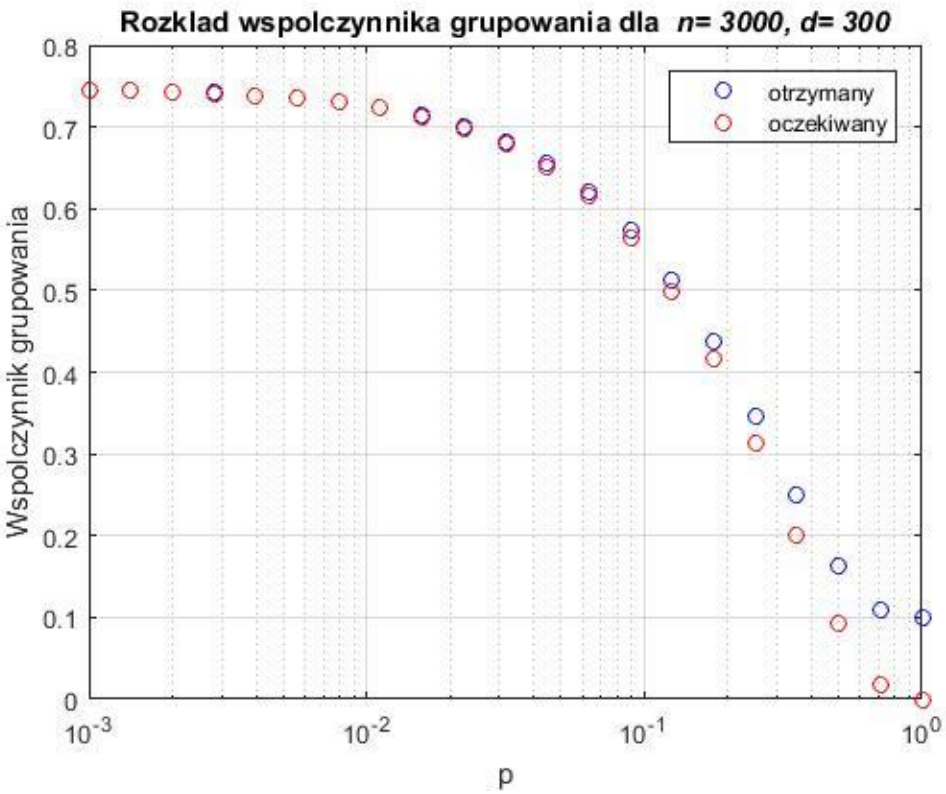
Testowanie

Generator był przetestowany na sieciach o różnej ilości wierzchołków: od 50 do 4000. Większej sieci nie udało się przetestować z powodu ograniczonych możliwości komputera, na którym był testowany generator. Również z powodu ograniczeń sprzętu i wybranego języka implementacji porównanie rozkładów stopni zrobione dla grafu o mniejszej ilości wierzchołków.

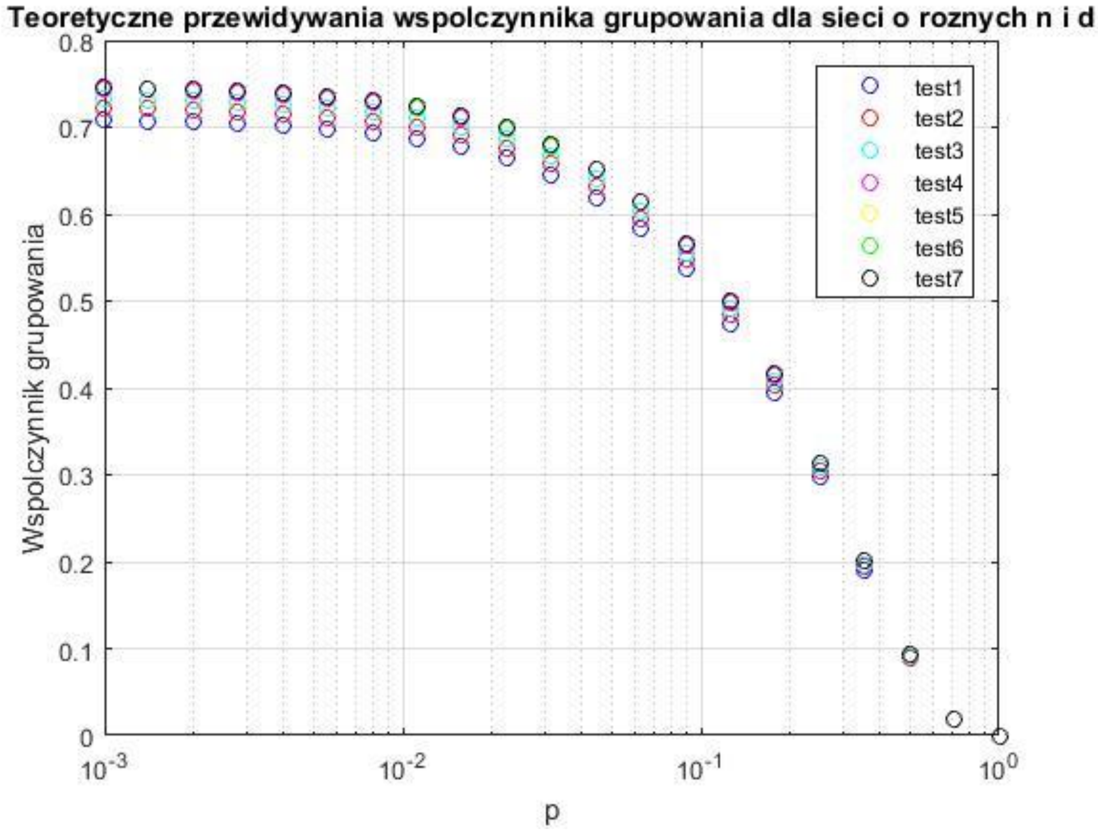
Przy testowaniu widzimy, że przewidywania teoretyczne współczynnika grupowania prawie się nie zmieniają, wielkość sieci nie ma dużego wpływu na wartość tego parametru (tab.3). Realny współczynnik grupowania sieci, obliczony przy pomocy generatora, różni się od przewidywań teoretycznych, wpływ na jego wartość ma parametr *n* i *d* (tab.2). Im bardziej zmiania się wartość parametru *p* od 0 do 1, tym bardziej większa różnica jest między teoretycznym i realnym parametrem (rys. 2, rys. 3).



rys. 2. Wykres współczynnika grupowania sieci o 4000 wierzchołkach

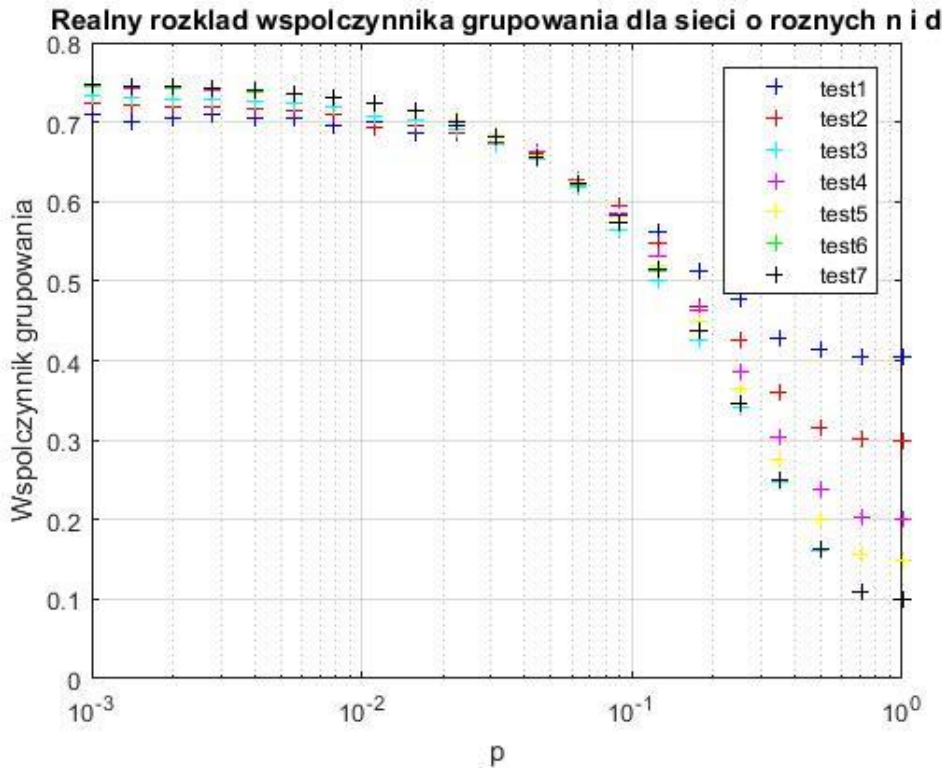


rys. 3. Wykres współczynnika grupowania sieci o 3000 wierzchołkach



rys. 4. Wykres współczynnika grupowania na podstawie testów z tabeli 3

Na rys. 4 można zobaczyć jak zmieniają się przewidywania teoretyczne współczynnika grupowania w zależności od parametrów *n* i *d*.



rys. 5. Wykres współczynnika grupowania na podstawie testów z tabeli 2

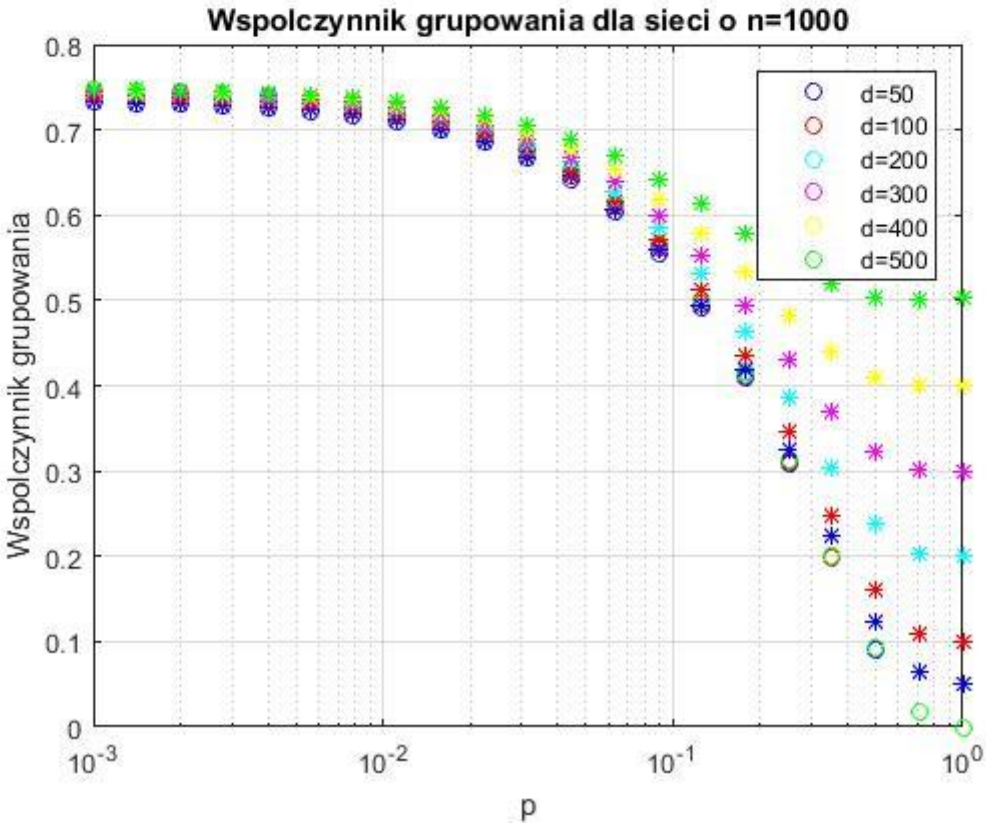
tab. 2. Realny współczynnik grupowania zaimplementowanej sieci

№ testu	n	d	p																				
			0.001	0.0014	0.002	0.0028	0.004	0.0056	0.0079	0.0112	0.0158	0.0224	0.0316	0.0447	0.0631	0.0891	0.1259	0.1778	0.2512	0.3548	0.5012	0.7079	1
1	50	20	0.7105	0.7001	0.7058	0.7099	0.7058	0.7048	0.6957	0.6995	0.6859	0.6954	0.6755	0.6606	0.6215	0.5833	0.5627	0.5128	0.4781	0.4273	0.4143	0.4039	0.4039
2	100	30	0.7232	0.7203	0.7198	0.7185	0.7165	0.7147	0.7106	0.6928	0.6965	0.6859	0.6754	0.6604	0.6239	0.5942	0.5470	0.4682	0.4256	0.3595	0.3156	0.3008	0.2980
3	500	50	0.7327	0.7319	0.7295	0.7293	0.7263	0.7243	0.7184	0.7085	0.7037	0.6906	0.6731	0.6537	0.6172	0.5645	0.5016	0.4258	0.3417	0.2475	0.1599	0.1089	0.099
4	1000	200	0.7441	0.7434	0.7420	0.7404	0.7380	0.7348	0.7299	0.7242	0.7137	0.7022	0.6831	0.6625	0.6270	0.5857	0.5312	0.4641	0.3853	0.3037	0.239	0.204	0.1998
5	2000	300	0.7454	0.7447	0.7432	0.7415	0.7392	0.7356	0.7309	0.7243	0.7150	0.7026	0.6836	0.6586	0.6245	0.5804	0.5201	0.4496	0.3650	0.2758	0.2005	0.1565	0.1498
6	3000	300	0.7454	0.7445	0.7434	0.7415	0.7386	0.7354	0.7304	0.7229	0.7142	0.7002	0.6821	0.6557	0.6214	0.5740	0.5134	0.4384	0.3470	0.25	0.1627	0.1099	0.0999
7	4000	400	0.7461	0.7449	0.7441	0.7420	0.7398	0.7362	0.7310	0.7244	0.7140	0.7006	0.6818	0.6566	0.6222	0.5746	0.5139	0.4374	0.3472	0.2499	0.1629	0.1098	0.0999

tab. 3. Przewidywania teoretyczne współczynnika grupowania

№ testu	n	d	p																				
			0.001	0.0014	0.002	0.0028	0.004	0.0056	0.0079	0.0112	0.0158	0.0224	0.0316	0.0447	0.0631	0.0891	0.1259	0.1778	0.2512	0.3548	0.5012	0.7079	1
1	50	20	0.7084	0.7075	0.7063	0.7045	0.7021	0.6986	0.6937	0.6869	0.6773	0.6639	0.6452	0.6195	0.5843	0.5370	0.4745	0.3949	0.2983	0.1908	0.0882	0.0177	0
2	100	30	0.7220	0.7211	0.7198	0.7180	0.7155	0.7120	0.7070	0.7	0.6903	0.6766	0.6576	0.6314	0.5955	0.5473	0.4836	0.4024	0.3040	0.1945	0.0899	0.0180	0
3	500	50	0.7325	0.7316	0.7303	0.7285	0.7260	0.7224	0.7173	0.7102	0.7003	0.6864	0.6672	0.6406	0.6042	0.5552	0.4907	0.4083	0.3085	0.1973	0.0912	0.0183	0
4	1000	200	0.744	0.7431	0.7418	0.7399	0.7374	0.7337	0.7286	0.7214	0.7113	0.6972	0.6777	0.6506	0.6137	0.564	0.4984	0.4147	0.3133	0.2004	0.0926	0.0186	0
5	2000	300	0.7453	0.7443	0.7430	0.7412	0.7386	0.7350	0.7298	0.7226	0.7125	0.6984	0.6788	0.6517	0.6147	0.5649	0.4992	0.4154	0.3139	0.2008	0.0928	0.0186	0
6	3000	300	0.7453	0.7443	0.7430	0.7412	0.7386	0.7350	0.7298	0.7226	0.7125	0.6984	0.6788	0.6517	0.6147	0.5649	0.4992	0.4154	0.3139	0.2008	0.0928	0.0186	0
7	4000	400	0.7459	0.7450	0.7437	0.7418	0.7392	0.7356	0.7304	0.7232	0.7131	0.699	0.6794	0.6523	0.6153	0.5654	0.4997	0.4158	0.3141	0.2009	0.0929	0.0186	0

Został także przeprowadzony test na grafie o stałej liczbie wierzchołków, zmieniany był jedynie początkowy stopień wierzchołków. Dla każdego grafu był obliczony współczynnik grupowania i jego teoretyczne przewidywania. Wykres jest przedstawiony na rys. 6.



rys. 6. Wykres współczynnika grupowania dla sieci o 1000 wierzchołkach

Graf na wykresie ma 1000 wierzchołków, zmiana parametru *d* od 50 do 500.

Kółkami jest przedstawiona zmiana przewidywań teoretycznych współczynnika grupowania, zaś gwiazdkami – zmiana współczynnika grupowania sieci, który jest w rzeczywistości.

Jak widzimy na wykresie, wartość współczynnika grupowania zależy od stopnia wierzchołka – im większy jest początkowy stopień wierzchołka, tym większy jest współczynnik grupowania.

Literatura

[1] Collective dynamics of ‘small-world’ networks. Duncan J. Watts & Steven H. Strogatz/ Department of Theoretical and Applied Mechanics, Kimball Hall, Cornell University, Ithaca, New York 14853, USA

[2] http://www.if.pw.edu.pl/~agatka/moodle/charakterystyki.html

[3] On the properties of small-world network models, A. Barrat and M. Weigt/ Eur. Phys. J. B 13, 547-560 (2000)

[4] J.Wojciechowski, K.Pieńkosz, Grafy i sieci, PWN, 2013, str. 382