

Modyfikacja modelu koewoluujących sieci społecznych

Chorowski Michał, Kowalski Filip, Ossowski Maciej, Thomson Ewa
Opiekun: mgr Tomasz Raducha

czerwiec 2019

1 Streszczenie

W poniższej pracy odtwarzamy model społecznych interakcji Axelroda na koewoluujących sieciach i porównujemy cechy stanów końcowych sieci dla dwóch wybranych modeli przełączania krawędzi. Jedną z metod jest dobrze znanym mechanizmem Barabasi-Albert, faworyzująca wierzchołki z dużą liczbą połączeń natomiast druga jest naszym autorskim pomysłem i opiera się na preferencji wspólnych sąsiadów. Dzięki temu powstaje sieć o innych własnościach.

2 Wstęp

W naszym świecie sieci są elementem, z którym spotykamy się na co dzień. Jednym z najbardziej oczywistych przykładów jest internet. Sieć tworzą serwery i komputery połączone ze sobą kablami. W człowieku również znajdują się sieci, jak na przykład krwiobieg czy sieć neuronów w mózgu. Ale sieci to nie tylko dosłowne połączenia. Poznając ludzi tworzymy sieci społeczne, dzięki temu już w czasach starożytnych wymieniano się informacjami i technologiami. Zazwyczaj używając internetu czy poznając nową osobę raczej nie zastanawiamy się nad tym jak można matematycznie modelować te zjawiska.

Pierwszą osobą która podjęła próbę takiego opisu był Euler, rozważając sieć mostów i tym samym zapoczątkował nową dziedzinę matematyki - teorię grafów. Grafy składają się z wierzchołków oraz połączeń pomiędzy nimi. Później również socjologowie zainteresowali się przedstawianiem relacji w społeczeństwie za pomocą sieci. Sieci mogą składać się z tysięcy, a nawet milionów elementów, połączenia z czasem mogą ulegać zmianom. Jest to złożony system i przez długi czas nie było odpowiedzi na to jak modelować taki ustrój.

W latach sześćdziesiątych nastąpił pierwszy przełom badania sieci złożonych, gdy dwóch węgierskich matematyków Paul Erdős i Alfred Rény postanowiło włączyć do teorii grafów rachunek prawdopodobieństwa. Dzięki temu nadano sieciom element przypadku który występuje w rzeczywistości. Mankamentem tak modelowanej sieci jest to, że średnio wszystkie wierzchołki mają taką samą

liczbę połączeń, a nie jest to cecha sieci rzeczywistych. W sieciach społecznych mamy osoby które są bardzo znane i obracające się w towarzystwie wielu różnych grup jak i osoby które mają stałą, mniejszą grupę znajomych.

Pierwszy zauważył ten fakt węgierski fizyk, Albert-Laszlo Barabasi. Wprowadził on modyfikacje aby zamiast losowego kreowania się nowych połączeń większe prawdopodobieństwo przyłączenia przypadało węzłom o większej liczbie już istniejących połączeń. Dzięki temu w symulacji sieci tworzą się tak zwane huby. Takie wierzchołki są bardzo istotnym elementem sieci. W rzeczywistości hubem mogłaby być elektrownia dostarczająca prąd do okolicznych domów. Jeśli jedna ulica zostanie odcięta od pozostałych nie będzie to duży problem, jednak gdy elektrownia przestanie działać to skutki mogą być dużo dalej idące. Model A w naszej pracy opiera się na modelu Barabasi-Albert (patrz punkt 3.2). Zainteresowanych pogłębieniem wiedzy odsyłamy do książki [4].

W naszej pracy szczególnie interesuje nas jak sieć społeczna ewoluuje, a więc połączenie modelu społecznych interakcji z koewoluującymi sieciami złożonymi (takimi gdzie połączenia ulegają zmianom). Model społecznych interakcji (znany jako model "the dissemination of culture" [1]) to koncept który został zdefiniowany przez Roberta Axerolda. Rozpoznał on cztery cechy które wpływają na stan końcowy sieci:

- F - liczba cech,
- q - liczba wartości jakie może przyjmować dana cecha,
- N - rozmiar sieci czyli liczba wierzchołków oraz
- liczba sąsiadów pierwszego rzędu (czyli bezpośrednich połączeń między wierzchołkami).

Model ten był rozwijany na wiele różnych sposobów i wiele ciekawych prac powstało na ten temat przypisując cechom różne interpretacje [6]. Jednak i ten model nie do końca opisuje rzeczywistą sytuację ponieważ zakłada, że tylko wierzchołki są dynamiczne (wartości cech ulegają zmianie). Nie opisuje to trafnie relacji w społeczeństwie ponieważ jak dobrze wiemy z czasem więzi pomiędzy osobami mogą wygasnąć lub pojawić się nowe. Maxi San Miguel wykazał, że wprowadzając model społecznych interakcji do sieci koewoluującej stan końcowy sieci jest rzeczywiście zależny również od jej struktury [5].

3 Model

Podobnie jak w oryginalnej pracy [6] generowany jest graf o N wierzchołkach, każdy reprezentuje jednego agenta. Każdy z i -tych agentów opisywany jest wektorem $\sigma_i = (\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,F})$. F jest liczbą cech opisujących każdego z agentów. Każda z cech może początkowo przyjąć jedną z q dyskretnych wartości, to znaczy $\sigma_{i,f} \in \{1, 2, \dots, q\}$, gdzie f przyjmuje wartości $f = 1, 2, \dots, F$. Jest zatem q^F możliwych stanów. Wartość q można interpretować na przykład jako możliwe preferencje, zainteresowania, poglądy, opinie i tym podobne na danej składowej

wektora $\sigma_{i,f}$ opisującej agenta. Wektor σ_i opisuje zatem opinie i charakterystykę kulturową danego agenta. Symulacja została wykonana w języku programowania python, a nasz kod jest dostępny na licencji GNU General Public License v3.0 na stronie: <https://github.com/Moribund7/AxelrodModel>.

3.1 Algorytm

Na początku każdemu z agentów przypisywane jest F cech, każda z wartości od 1 do q ma równe prawdopodobieństwo wylosowania. Następnie algorytm przebiega następująco, tak jak w pracy [6]:

1. Wybierz aktywny wierzchołek i i jego sąsiada j , czyli inny wierzchołek, który jest połączony z i .
2. Porównaj wektory σ_i i σ_j i określ liczbę identycznych cech m , to znaczy takich, dla których $\sigma_{i,f} = \sigma_{j,f}$.
 - (a) Jeśli wszystkie cechy są równe $m = F$, nic się nie dzieje.
 - (b) Jeśli żadna z cech nie jest taka sama $m = 0$, rozłącz krawędź łączącą (i, j) , wybierz nowy wierzchołek l i utwórz z nim krawędź (i, l) .
 - (c) W pozostałych przypadkach, z prawdopodobieństwem m/F zachodzi interakcja między agentami i zostaje losowo wybrana jedna z cech, która nie jest wspólna f' spośród $F - m$. Wierzchołek i przyjmuje tę wylosowaną wartość, czyli $\sigma_{i,f'} \rightarrow \sigma_{j,f'}$.
3. Idź do następnego kroku czasowego.

W ogólności punkt 2b algorytmu może być realizowany na różne sposoby. W tej pracy rozpatrzyliśmy dwa sposoby. Pierwszy replikuje jeden z modeli w oryginalnej pracy [6] i stanowi punkt odniesienia do drugiego sposobu, który jest naszą autorską modyfikacją. Oba modele zostały opisane poniżej.

3.2 Modele przełączania krawędzi

Prawdopodobieństwo wyboru nowego wierzchołka j stopnia k_j (liczba krawędzi łączących wierzchołek z innymi) dla aktywnego wierzchołka i realizujemy na następujące sposoby.

1. Model A z preferencją wysokiego stopnia.

- $$P_A(j) = \frac{k_j}{\sum_{a=1}^N k_a}$$

2. Model F z preferencją wierzchołków odległych o dwie krawędzie mających najwięcej wspólnych sąsiadów z wierzchołkiem i .

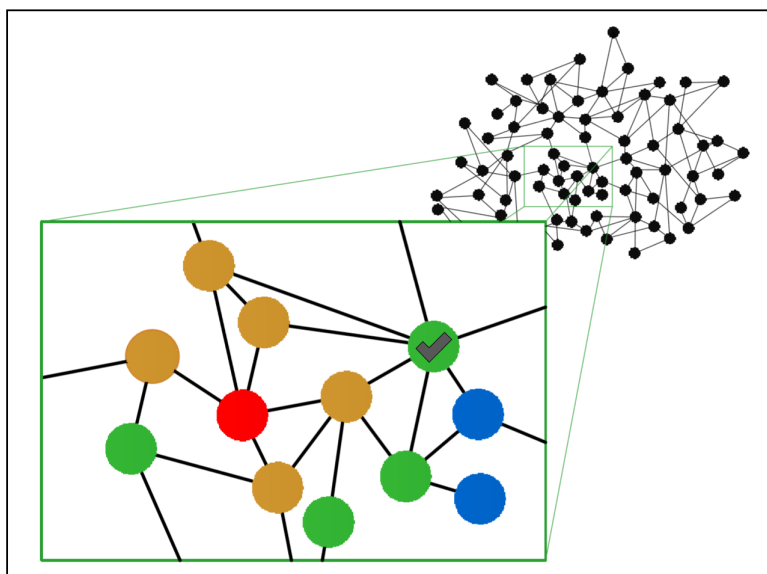
3.3 Zasada działania modelu A

Model A faworyzuje wierzchołki z większą liczbą krawędzi. Zjawisko takie jest typowe na przykład dla cytowań w nauce, ale niekoniecznie oddaje on mikroskopowy obraz typowych, codziennych interakcji [6]. Dlatego w oryginalnej pracy zaproponowany został model D, w którym nowy sąsiad jest wybierany z rozkładu jednorodnego na zbiorze sąsiadów najbliższych sąsiadów aktywnego wierzchołka. W interpretacji interakcji społecznych oznacza to, że w modelu D aktywny wierzchołek znajduje nowych znajomych spośród znajomych swoich znajomych z jednakowym prawdopodobieństwem. Nasza modyfikacja w modelu F inspirowana była działaniem portalu Facebook, którego użytkownicy tworzą ogromną sieć powiązań. Na portalu Facebook użytkownikom proponowani są nowi znajomi. Obok każdego proponowanego znajomego wyświetlona jest liczba wspólnych znajomych, których użytkownik już ma z proponowanym kontem. Spodziewając się, że prawdopodobieństwo zawarcia znajomości z użytkownikiem, z którym mamy więcej wspólnych znajomych jest większe niż z takim, z którym wspólnych znajomych mamy mniej, zaproponowaliśmy model F, który może lepiej uchwycić tę sytuację. Nasza intuicja może nie być bezpodstawna, gdyż procedura szukania nowych znajomych wśród znajomych osób znajomych jest centralna dla portali takich jak Facebook [2], choć dane dotyczące skuteczności tej strategii nie są powszechnie dostępne [3]. Nie wiadomo zatem jak istotnie w praktyce liczba wspólnych znajomych wpływa na prawdopodobieństwo zawarcia nowej znajomości.

3.4 Zasada działania modelu F

Gdy wybrany jest aktywny wierzchołek i oraz znajdujemy się w punkcie 2b algorytmu, w modelu F prawdopodobieństwo stworzenia krawędzi z nowym wierzchołkiem (i, l) obliczane jest w następujący sposób. Na początek stwórz listę L . Następnie:

1. Wybierz wierzchołek j oddalony od wierzchołka aktywnego i o dwie krawędzie, czyli sąsiada jednego z najbliższych sąsiadów (znajomy znajomego)
2. Policz ile jest krawędzi łączących j z dowolnym z bezpośrednich sąsiadów i i za każdą taką krawędź dodaj wierzchołek j do listy L (liczymy wspólnych znajomych i z j).
3. Powtarzaj od kroku 1 dla kolejnego wierzchołka oddalonego o dwie krawędzie od i , aż nie zostaną wybrane wszystkie wierzchołki oddalone o dwie krawędzie od i .



Rysunek 1: Schematyczne przedstawienie działania modelu F. Czerwony kolor oznacza aktywny wierzchołek, złoty sąsiadów pierwszego rzędu, zielony sąsiadów drugiego rzędu, a niebieski trzeciego rzędu. Wierzchołek o kolorze zielonym z szarym znaczkiem jest sąsiadem drugiego rzędu o największej ilości wspólnych sąsiadów pierwszego rzędu z wierzchołkiem aktywnym (czerwonym).

Prawdopodobieństwo wybrania wierzchołka j na nowego sąsiada aktywnego wierzchołka i to liczba krawędzi łączących j z dowolnym z bezpośrednich sąsiadów i dzielona przez liczbę elementów listy L . Ponieważ w kroku 2 dany wierzchołek jest policzony tyle razy ile miał wspólnych znajomych, czyli mógł być policzony kilkakrotnie, zatem prawdopodobieństwo połączenia z takim wierzchołkiem jest większe. Oznacza to, zgodnie z naszym modelem, że prawdopodobieństwo połączenia jest tym większe im więcej dane wierzchołki mają wspólnych znajomych.

3.5 Zakończenie symulacji

Symulacja jest sprawdzana co 10,000 kroków czasowych symulacji i zostaje przerwana jeśli jeden z poniższych warunków zostaje spełniony:

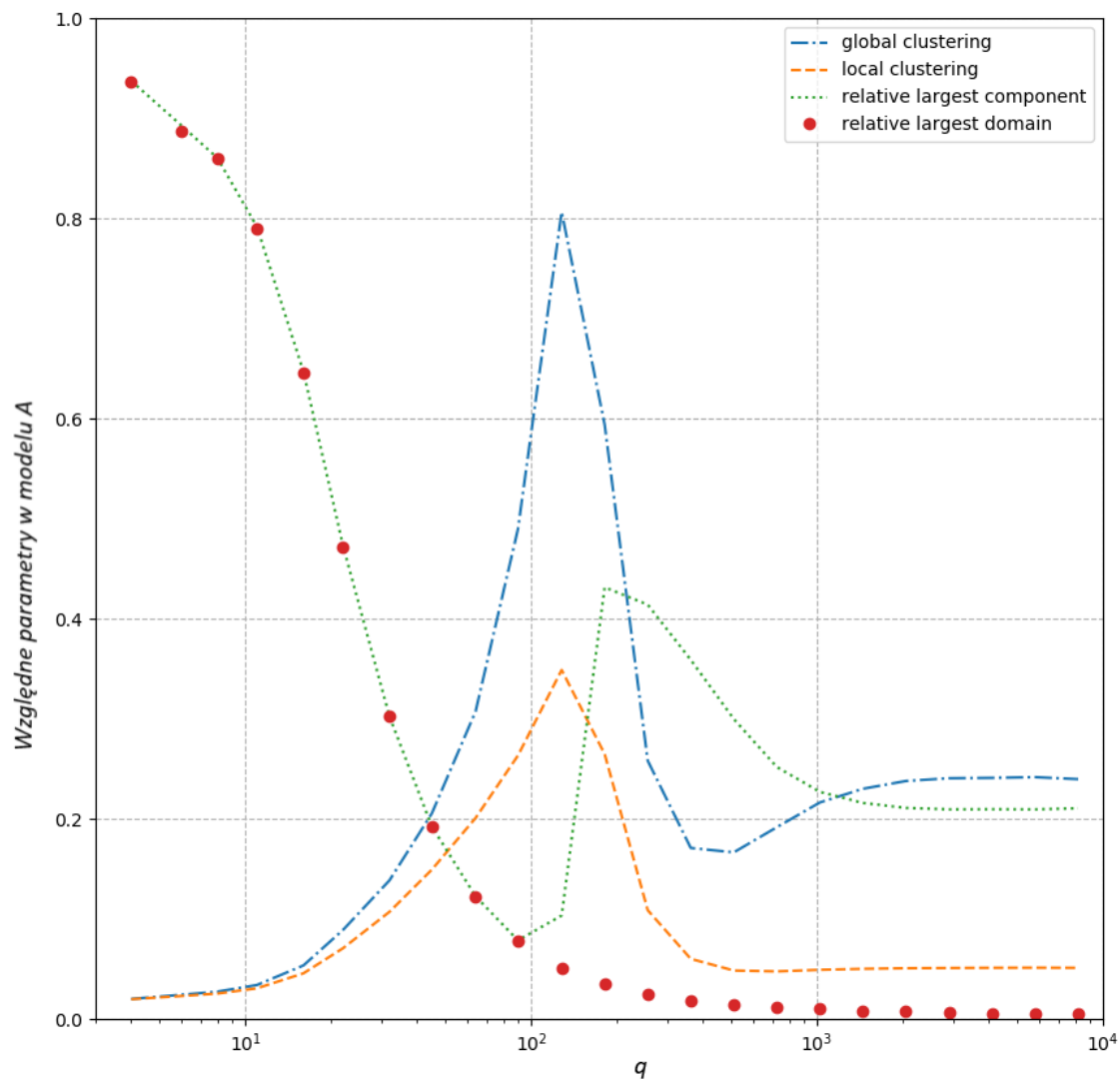
- liczba aktywnych połączeń, czyli takich, w których wektory σ nie są sobie równe, wynosi zero,
- liczba aktywnych połączeń w ostatnich 20-stu sprawdzeniach (200,000 kroków czasowych) nie uległa zmianie, zakładamy bowiem, że już się ona nie zmieni.

4 Wyniki

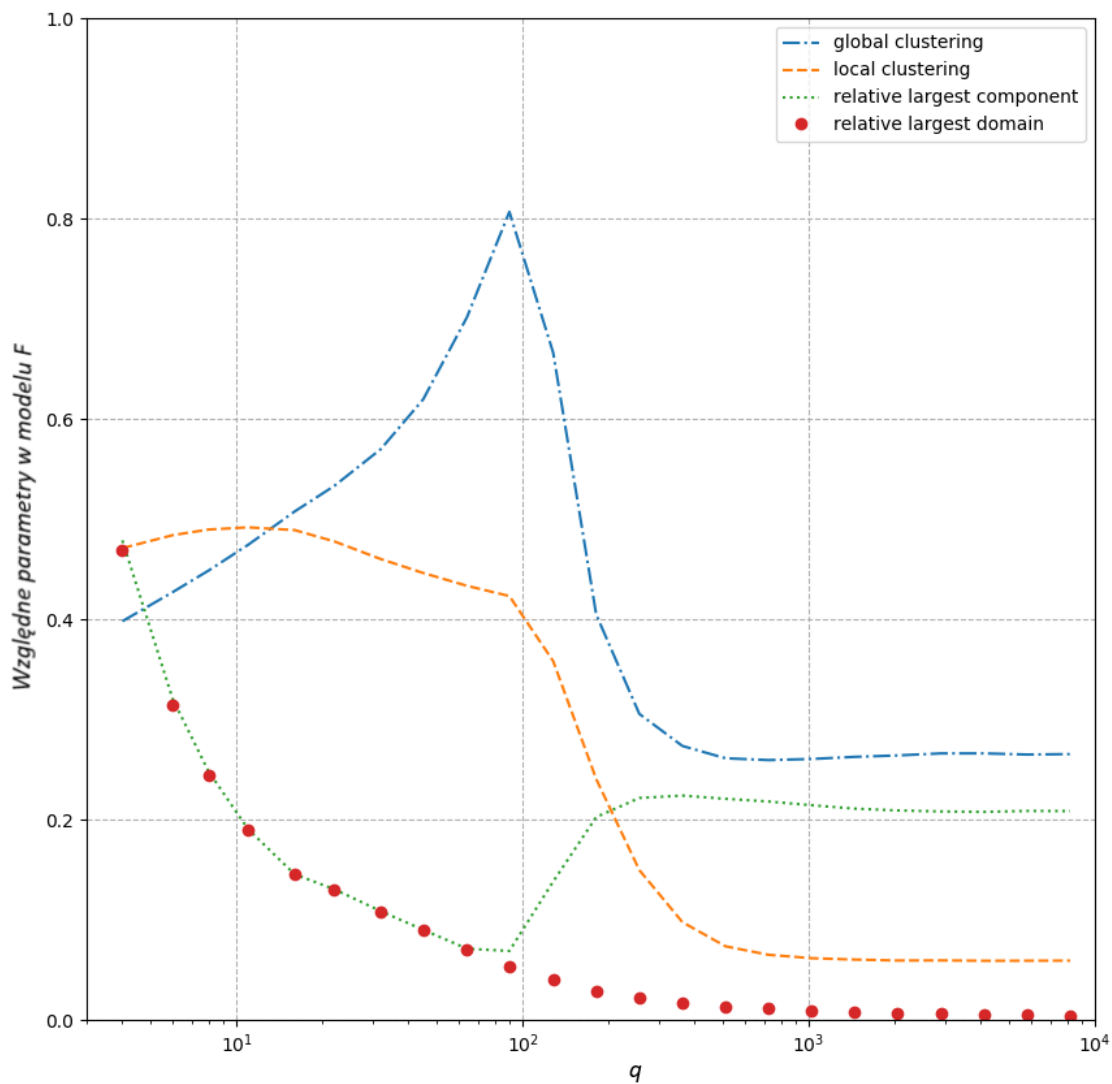
Aby określić jak wprowadzona przez nas modyfikacja wpływa na końcowy stan sieci obliczyliśmy średnie wartości następujących wielkości:

- Globalny współczynnik klastrowania (global clustering coefficient) - jest określony na trójkach wierzchołków. Zamkniętym trójkątem jest grupa trzech wierzchołków połączonych trzema krawędziami natomiast otwartym, gdy są połączone dwoma. Globalny współczynnik klastrowania to stosunek zamkniętych trójkątów do wszystkich (otwartych i zamkniętych).
- Lokalny współczynnik klastrowania znany również jako współczynnik gromowania wierzchołka (local clustering coefficient) - liczba istniejących połączeń pomiędzy sąsiadami danego wierzchołka podzielona przez liczbę możliwych połączeń. Przyjmujemy, że jeżeli dany wierzchołek ma mniej niż dwóch sąsiadów to jego lokalny współczynnik klastrowania równy jest zero, ale nadal jest on brany pod uwagę w obliczeniach.
- Komponent s - jest to podgraf czyli grupa wierzchołków które są połączone ze sobą ścieżką. Komponentem może być pojedynczy wierzchołek jak i cała sieć może być jednym komponentem, gdy wszystkie wierzchołki są połączone.
- Względny największy komponent S (relative largest component) - największy komponent podzielony przez liczbę wierzchołków N .
- Domena d - grupa wierzchołków posiadająca taki sam wektor cech σ które są połączone ścieżką i każdy wierzchołek na ścieżce również należy do domeny.
- Względna największa domena (relative largest domain) - największa domena D , czyli najliczniejsza pod względem wierzchołków podzielona przez całkowitą liczbę wierzchołków N .

Nasze wyniki przedstawione są na rysunkach na dwóch kolejnych stronach. Wykres na Rysunku 2 jest dla sieci z modelu A, natomiast na Rysunku 3 są wyniki dla modelu F. Na osi poziomej znajduje się wartość q gdzie liczba cech w symulacji wynosiła $F=3$. Na osi pionowej zaznaczona jest bezwymiarowa wielkość badanych parametrów w odniesieniu do wielkości sieci N . Każdy punkt odpowiada uśrednionej wartości z 400 przebiegów symulacji dla każdej wartości q . W symulacji zaczynamy od grafu losowego o średnim stopniu $\langle k \rangle = 4$, zatem w każdej chwili istnieje 1000 krawędzi.



Rysunek 2: Wyniki symulacji dla modelu A. Na osi pionowej znajdują się względne parametry w modelu A. Na osi poziomej znajduje się q , czyli liczba możliwych wartości dla składowych wektora σ . Kolorem niebieskim zaznaczono globalny współczynnik klastrowania, pomarańczowym lokalny współczynnik klastrowania, zielonym średni względny największy komponent, czerwonymi kropkami największą domenę. Każdy punkt odpowiada uśrednionej wartości z 400 przebiegów symulacji i sieci o $N = 500$ wierzchołkach. W symulacji zaczynamy od grafu losowego o średnim stopniu $\langle k \rangle = 4$, zatem w każdej chwili istnieje 1000 krawędzi.



Rysunek 3: Wyniki symulacji dla modelu F. Na osi pionowej znajdują się względne parametry w modelu F. Na osi poziomej znajduje się q , czyli liczba możliwych wartości dla składowych wektora σ . Kolorem niebieskim zaznaczono globalny współczynnik klastrowania, pomarańczowym lokalny współczynnik klastrowania, zielonym średni względny największy komponent, czerwonymi kropkami największą domenę. Każdy punkt odpowiada uśrednionej wartości z 400 przebiegów symulacji i sieci o $N = 500$ wierzchołkach. W symulacji zaczynamy od grafu losowego o średnim stopniu $\langle k \rangle = 4$, zatem w każdej chwili istnieje 1000 krawędzi.

5 Dyskusja

Wyniki dla modelu A, jak widać na Rysunku 2 otrzymaliśmy zbliżone do wyników w pracy [6]. Jedyna widoczna różnica występuje dla lokalnego klastrowania i wynika ona być może z różnego sposobu obliczeń. W naszym programie przyjęliśmy, że gdy dany wierzchołek ma mniej niż dwóch sąsiadów to jego lokalny współczynnik klastrowania jest równy zero, ale nadal jest brany pod uwagę przy obliczeniach.

Na Rysunku 2 zauważyć można, że w pierwszej fazie sieć składa się praktycznie wyłącznie z jednego dużego komponentu i jest on równy największej domenie. Druga faza charakteryzuje się rozdzieleniem na wiele mniejszych komponentów. W fazie trzeciej największy względny komponent S/N jest nadal większy niż w fazie drugiej, ale rozmiar względnej największej domeny D/N maleje wraz ze wzrostem q .

W zmodyfikowanym modelu F na Rysunku 3 zauważyć można kilka różnic w porównaniu do modelu A na Rysunku 2 i modelu D opisanego w pracy [6]. Po pierwsze na Rysunku 3 nie widać wyraźnej granicy między fazą I a II patrząc na względny rozmiar największego komponentu S/N . Wynika to z faktu, że minimalne badane $q = 4$. Spodziewamy się, że dla mniejszego q wartość S/N będzie dużo wyższa. Oznacza to, że doszło do przesunięcia q , przy którym następuje pierwsze przejście fazowe. Po drugie, już dla niewielkich wartości q zaobserwować można lokalne i globalne klastrowanie na poziomie około 0.5. Widoczne jest natomiast przejście fazowe w okolicach $q = 10^2$. W okolicach tej wartości największy komponent gwałtownie rośnie i będzie się dalej utrzymywał na stabilnym poziomie, natomiast największa domena stale maleje. Faza III jest podobna do tej w modelu D, kiedy wszystkie mierzone wielkości nie zmieniają się już.

6 Podsumowanie

Jak widać, wprowadzona przez nas modyfikacja znacząco zmienia zachowanie się sieci. Ciekawym zjawiskiem jest to, że przebieg lokalnego klastrowania jest zupełnie inny niż w modelu opartym na preferowaniu węzłów o dużej liczbie wierzchołków. Wyraźne różnice występują szczególnie dla małych q . W obu modelach możemy zaobserwować 3 fazy jednak oba przejścia fazowe w naszym modelu występują dla mniejszych wartości q . W następnym kroku myślimy, że warto byłoby odtworzyć sieć połączeń z danych rzeczywistych z Facebooka i porównać te parametry. Wydaje nam się jednak, że otrzymanie takich danych (oczywiście anonimowych) nie jest na ten moment możliwe.

Literatura

- [1] Robert Axelrod. The dissemination of culture a model with local convergence and global polarization. *Conflict Resolution*, 41(2):203–226, 1997.

- [2] Taina Bucher. The friendship assemblage: Investigating programmed sociality on facebook. *Television & New Media*, 14(6):479–493, 2013.
- [3] Jilin Chen, Werner Geyer, Casey Dugan, Michael Muller, and Ido Guy. Make new friends, but keep the old: recommending people on social networking sites. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 201–210. ACM, 2009.
- [4] Agata Fronczak and Piotr Fronczak. *Świat sieci złożonych, od fizyki do internetu*. Wydawnictwo Naukowe PWN, 2009.
- [5] K. Klemm, V.M. Eguiluz, R. Toral, and M. San Miguel. Role of dimensionality in axerold’s model for the dissemination of culture. *Physica A*, 327(1):1–5, 2003.
- [6] Tomasz Raducha and Tomasz Gubiec. Coevolving complex networks in the model of social interactions. *Physica A: Statistical Mechanics and its Applications*, 471:427–435, 2017.