



Politechnika Warszawska
Wydział Elektroniki i Technik Informacyjnych

Grafy i Sieci (GIS)

TEORIA GRAFÓW W MODELOWANIU EPIDEMII

SPRAWOZDANIE 2

Michał Dobrzański
mdobrzezan@mion.elka.pw.edu.pl

Maciej Janusz Krajsman
M.Krajsman@stud.elka.pw.edu.pl

Opiekun projektu:
mgr inż. Łukasz Błaszczuk

5 grudnia 2016

Spis treści

1	Szczegółowy opis merytoryczny zadania	4
2	Opis algorytmu generowania grafów losowych według modelu sieci Barabásiego-Albert	4
3	Szczegółowy opis modelu SIS rozwoju epidemii	5
3.1	Założenia modelu SIS	5
3.2	Warunki początkowe	6
3.3	Równania różniczkowe opisujące model	7
4	Opis planu zastosowania grafów losowych do omawianego zagadnienia . .	9
5	Podstawowe założenia implementowanego programu oraz projekt testów .	10
Bibliografia		12

1 Szczegółowy opis merytoryczny zadania

Zadaniem projektowym jest **zaimplementowanie modelu SIS** rozwoju epidemii, zbadanie jego właściwości na grafach modelujących populację ludzi, w których wierzchołek odpowiada konkretnej osobie. Należy również porównać otrzymane wyniki z modelem ciągłym dla epidemii opisanym równaniami różniczkowymi.

Zadanie będzie polegało na **wygenerowaniu odpowiednich grafów losowych**, a następnie na uruchomieniu **algorytmu propagacji epidemii** (zgodnego z modelem SIS) dla wierzchołków tych grafów.

Otrzymane wyniki zostaną zestawione z wynikami otrzymanymi za pomocą równań różniczkowych. Dla przejrzystości utworzone zostaną wykresy porównujące oba podejścia. Dobór prezentowanych parametrów dla osi wykresów zostanie określony w trakcie tworzenia projektu.

2 Opis algorytmu generowania grafów losowych według modelu sieci Barabásiego-Albert

Do zrealizowania projektu potrzebny będzie algorytm generujący grafy losowe według **modelu sieci Barabásiego-Albert**. Utworzone sieci dzięki temu modelowi są nazywane sieciami przypadkowymi ewoluującymi.

Zaproponowana przez autorów **ewolucja sieci** rozumiana jako zmiany struktury sieci w kolejnych odstępach czasowych modelowana jest poprzez dołączanie nowych węzłów do istniejącej już sieci.

Procedura konstruowania sieci Barabásiego-Albert obejmuje następujące kroki:

1. Na początkowym etapie ewolucji (czyli w chwili $t = 0$) siecią nazywamy graf pełny (całkowicie połączony klaster węzłów) o rozmiarze: $m_0 \geq 1$. W następnych krokach czasowych $t = 1, 2, 3, \dots$ do sieci dodawane są nowe węzły (jeden węzeł na jeden krok), które tworzą odpowiednio $m \leq m_0$ ($m = \text{const}$) połączeń (czyli krawędzi) do istniejących już wierzchołków sieci.
2. Proces dodawania wierzchołków realizuje **regułę preferencyjnego dołączania**, która mówi o tym, że prawdopodobieństwo, że nowy wierzchołek utworzy połącze-

nie do jednego ze starszych wierzchołków jest wprost proporcjonalne do stopnia wierzchołka starszego.

3. Wzrost sieci kończony jest w dowolnej chwili t . W momencie zakończenia wzrostu sieć ma:

$$N = t + m_0 \approx t$$

wierzchołków (węzłów) oraz

$$E = mt + \binom{m_0}{2} \approx mt$$

krawędzi.

3 Szczegółowy opis modelu SIS rozwoju epidemii

Model SIS opiera się na zestawie równań różniczkowych, które opisują **rozprzestrzanie się chorób zakaźnych**. Służy do określania, czy dana choroba **zaniknie**, czy **ustali się** na konkretnym poziomie. Jest najbardziej ogólnym modelem matematycznym tego zjawiska — nie uwzględnia np. przypadków nosicielstwa i uodpornienia, istnienia osobników w fazie utajonej choroby czy urodzeń chorych. Stan populacji w tym modelu opisany jest **tylko jedną zmienną (I)**, która reprezentuje liczbę osobników zainfekowanych.

3.1 Założenia modelu SIS

- Jednostki mogą znajdować się w jednym z dwóch stanów:

S (*susceptible* — *ang. podatni*) — jednostki są zdrowe, podatne na zakażenie.

I (*infected* — *ang. zainfekowani*) — jednostki są chore, mogą zakażać zdrowe jednostki.

S(t) — liczba osobników podatnych (zdrowych), tj. liczba węzłów w stanie S w danej chwili czasowej t .

I(t) — liczba jednostek zainfekowanych, tj. liczba węzłów w stanie I w danej chwili czasowej t .

- Do zakażenia (zmiany stanu węzła z S na I może dojść na skutek kontaktu jednostki zdrowej i chorej.
- N — liczebność populacji, dla której $N = S(t) + I(t)$, Liczebność populacji jest stała, tj. $N = \text{const}$.
- β — prawdopodobieństwo, że w pojedynczym kroku czasowym (dt) zdrowy osobnik zakaży się od chorego sąsiada.
- γ — prawdopodobieństwo, że w pojedynczym kroku czasowym (dt) chory osobnik wyzdrowieje, tj. zmieni stan $I \rightarrow S$.
- λ — parametr określający **tempo rozprzestrzeniania się epidemii**. Jest to stosunek $\lambda = \beta/\gamma$.
- λ_c — **wartość progowa tempa rozprzestrzeniania się epidemii**. Gdy tempo przekroczy tę wartość krytyczną, wówczas epidemia ma szansę stać się powszechną, a badana choroba nabrać charakteru endemicznego. Dla symulacji w czasie nieograniczonym oznacza to endemię.

3.2 Warunki początkowe

Dla sieci losowych oraz bezskalowych (BA) dzięki modelowi SIS można opisać propagację epidemii za pomocą równania różniczkowego określającego tempo zmiany w czasie liczby zakażonych węzłów o zadanym stopniu k . W tym celu należy określić **warunki początkowe** epidemii, czyli w chwili $t = 0$. Określa się:

- $P(k)$ — rozkład stopni wierzchołków otrzymany po wygenerowaniu grafu BA dla wszystkich wartości k .
- β, γ — określone prawdopodobieństwa odpowiednio: zdrowy zostanie zakażony przez chorego, chory wyzdrowieje.
- $I_k(0)$ — początkowa liczba zainfekowanych węzłów o stopniu k .
- $S_k(0)$ — początkowa liczba podatnych węzłów o tym samym stopniu k .

Pytanie, jakie zadaje się przy modelowaniu epidemii najczęściej brzmi: Czy dla zadanych wartości parametrów β i λ oraz dla zadanej początkowej liczby zainfekowanych osobników $i_0(0)$ infekcja rozprzestrzeni się czy nie?

Ponadto dla sieci bezskalowych szuka się wartości parametru λ_c , czyli progu od którego epidemia może stać się powszechna w danej populacji.

3.3 Równania różniczkowe opisujące model

Oprócz wcześniej wymienionych parametrów dla modelu SIS w sieciach bezskalowych określa się prawdopodobieństwo, że dowolna krawędź grafu prowadzi do węzła, który przechowuje stan o wartości „chory”.

$$Q_I = \sum_k Q(k) i_k \quad (1)$$

Gdzie:

k — stopień wierzchołka.

Q_I — prawdopodobieństwo, że dowolna krawędź grafu prowadzi do węzła, który przechowuje stan o wartości „chory”

$Q(k)$ — prawdopodobieństwo, że dowolna krawędź grafu prowadzi do węzła o stopniu k , który przechowuje stan o wartości „chory”.

i_k — prawdopodobieństwo, że węzeł o stopniu k przechowuje stan o wartości „chory”.

Wartość $Q(k)$ można dla danej sieci powiązań dla danej populacji wyliczyć z następującego wzoru:

$$Q(k) = \frac{k}{\langle k \rangle} P(k) \quad (2)$$

Gdzie:

k — stopień wierzchołka.

$\langle k \rangle$ — gęstość sieci. Średni stopień losowo wybranego wierzchołka.

W przypadku sieci BA o tworzeniu połączeń w sieci nie decyduje żadne z góry narzucone prawo, jej **struktura zależy od indywidualnych decyzji jej elementów** (kto kogo poznał, kto się z kim widuje). W rezultacie otrzymujemy **potęgowy rozkład stopni wierzchołków**, w którym w bliskim sąsiedztwie mogą występować wierzchołki o ogromnej liczbie incydentnych krawędzi oraz te niemalże odizolowane od reszty sieci. Ze względu

na potęgowy rozkład stopni wierzchołków, wyniki byłyby niewłaściwe. Równania korzystające ze średniego stopnia $\langle k \rangle$ ilustrują dynamikę układu zawierającego same wierzchołki o średnim stopniu. W opisie dynamiki sieci BA należy więc stosować równania dotyczące wierzchołków o **zadanym stopniu k** .

Tempo zmiany osobników zainfekowanych (węzłów o zadanym stopniu k) $I_k(t)$ opisuje się następująco:

$$\frac{dI_k(t)}{dt} = [\beta k Q_I] S_k(t) - \gamma I_k(t) \quad (3)$$

Można również wyliczać tempo zmiany prawdopodobieństwa, że konkretny osobnik (węzeł o zadanym stopniu k) jest zainfekowany:

$$\frac{di_k(t)}{dt} = [\beta k Q_I] s_k(t) - \gamma i_k(t); \quad i_k(t) + s_k(t) = 1 \quad (4)$$

Szczególnie interesujący jest przypadek graniczny, dla $i_k = \lim_{t \rightarrow \infty} i_k(t)$:

$$i_k = \frac{\lambda k Q_I}{1 + \lambda k Q_I} \quad (5)$$

Zgodnie z zapisem w sekcji 3.1, parametr $\lambda = \beta/\gamma$ określa tempo rozprzestrzeniania się epidemii. W powyższym równaniu (5) określone zostało prawdopodobieństwo że węzeł o stopniu k jest w stanie I (jest zakażony) w czasie nieskończonym, czyli po zakończeniu symulacji. Wyzerowanie i_k zależy od Q_I , ponieważ k jest dodatnią liczbą naturalną, a $\lambda > 0$. By wykonywać obliczenia dla wszystkich stopni k , należy zsumować obliczenia dla każdego z nich. Z wzorów (6), (2) i (5) wynika, że:

$$Q_I = \sum_k Q(k) i_k = 1 - \frac{1}{\langle k \rangle} \sum_k P(k) \frac{k}{1 + \lambda k Q_I} = f(Q_I) \quad (6)$$

Gdzie:

k — stopień wierzchołka.

$f(Q_I)$ — funkcja rozkładu prawdopodobieństwa przechowywania stanu „chory” przez węzły o różnych stopniach.

Powyższe równanie może mieć jedno lub dwa rozwiązania. Jednym jest zawsze $Q_I = 0$ (dla $i_k = 0$). Obecność drugiego rozwiązania świadczy o tym, że losowo wybrana krawędź

proceedzi do zainfekowanego wierzchołka. Epidemia może się swobodnie rozwijać, więc $\lambda > \lambda_c$. Dzięki temu, warunek:

$$\left. \frac{df(Q_I)}{dQ_I} \right|_{Q_I=0} > 1 \quad (7)$$

umożliwia wyznaczenie λ_c . By funkcja (6) spełniała warunek (7):

$$\left. \frac{df(Q_I)}{dQ_I} \right|_{Q_I=0} = \frac{1}{\langle k \rangle} \sum_k P(k) \frac{\lambda k^2}{(1 + \lambda k Q_I)^2} \Big|_{Q_I=0} = \frac{\lambda}{\langle k \rangle} \sum_k k^2 P(k) = \lambda \frac{k^2}{k} > 1 \quad (8)$$

Więc:

$$\lambda_c = \frac{\langle k \rangle}{\langle k^2 \rangle} \quad (9)$$

Okazuje się, że w sieciach bezskalowych wartość parametru λ_c jest bardzo zbliżona do zera. W praktyce powinno to oznaczać, że powstrzymanie rozwoju epidemii jest prawie niewykonalne.

4 Opis planu zastosowania grafów losowych do omawianego zagadnienia

Wygenerowane grafy losowe według modelu Barabasiiego-Arbert posłużą za model badanej populacji. **Wierzchołkami** będziemy reprezentowali poszczególne osoby w danej populacji. **Krawędzie** będą przedstawiały natomiast relacje pomiędzy tymi osobami. Określamy, że dwa wierzchołki są sąsiednie, gdy mają wspólną krawędź grafu.

Siecią losową, bezskalową (w zadanej chwili t utworzoną według modelu BA) będziemy nazywać określoną populację o ustalonej liczności, która zostanie poddana modelowaniu zjawiska propagacji epidemii.

Wierzchołki grafu będą przechowywały informacje o stanie osoby — czy jest ona zdrowa (podatna), czy chora (zakażająca) według modelu SIS. Zostanie zastosowana reprezentacja grafu za pomocą listy wierzchołków wraz z odpowiadającymi im listami sąsiedztwa. Badanie rozwoju epidemii będzie bazowało na tej reprezentacji opisującej stan, w jakim się dany graf znajduje.

5 Podstawowe założenia implementowanego programu oraz projekt testów

Program zostanie zaimplementowany w środowisku Python. Do jego realizacji użyje się bibliotek niezbędne do rysowania wykresów oraz grafów.

Struktura grafu będzie reprezentowana za pomocą listy sąsiedztwa. Dodatkowo w niej dla każdego wierzchołka umieści się dodatkową informację o jego aktualnym stanie (S lub I według modelu SIS). W środowisku Python będzie to słownik tworzony za pomocą konstruktora *dict()*. Badanie algorytmu propagacji epidemii sprowadzi się do przechodzenia po listach sąsiedztwa i zmianach stanów dla wierzchołków.

W implementowanym programie przeprowadzone zostaną następujące testy:

1. Sprawdzenie poprawności algorytmu generującego grafy losowe według modelu sieci Barabasiiego-Albert:

Zostaną wygenerowane sieci losowe BA (o różnych parametrach m_0 , m oraz t). Ich rozkłady stopni wierzchołków zostaną przedstawione na wykresie. Prawidłowo utworzone sieci według modelu BA powinny charakteryzować się w przybliżeniu **potęgowym rozkładem stopnia wierzchołka**. Ten rozkład przybiera postać liniowego przebiegu na wykresie w skali podwójnie logarytmicznej. Na tym wykresie wartość nachylenia krzywej przekłada się na wartość wykładnika dla rozkładu potęgowego. W programie utworzony zostanie taki wykres i przedstawione na nim będzie kilka przykładowych rozkładów dla różnych sieci losowych BA.

2. Przeprowadzenie symulacji rozwoju epidemii według modelu SIS dla różnych parametrów:

Następnie dla wcześniej wygenerowanych sieci losowych przeprowadzona zostanie symulacja rozwoju epidemii. Dla każdej sieci utworzone będzie kilka różnych warunków początkowych (odpowiednia liczba wierzchołków w stanie S i I w chwili t_0 , różne parametry γ , β). Następnie dla kolejnych chwil $t = 1, 2, 3, \dots$ będzie badany rozwój epidemii. Najistotniejszy jest wynik dla $t \rightarrow \infty$, ponieważ opisuje stan końcowy układu. Wyniki porównane zostaną ze spodziewanym rezultatem otrzymanym z równań opisujących model SIS. Sprawdzone zostanie również, jak model zachowuje się dla $\lambda > \lambda_c$ oraz $\lambda < \lambda_c$ (w tym celu graf nie może być za mały,

a wartości λ_c i tak będą bardzo niewielkie, ze względu na potęgowy rozkład stopni wierzchołków grafu).

Bibliografia

- [1] Fronczak Agata, Fronczak Piotr: „*Świat sieci złożonych*”,
Wyd. I, Warszawa, wyd. PWN, 2009, ISBN: 978-83-01-15987-0.
- [2] Fronczak Agata, Fronczak Piotr: „*Epidemie w sieciach złożonych*” [online],
28.10.2010, <http://www.if.pw.edu.pl/~agatka/moodle/epidemie.html> ,
dostęp: 20.11.2016r.