Grafy i sieci:
Generator sieci bezskalowej II (model Barabasiego-Albert)

Eryk Warchulski
Kanstantsin Padmostka
Prowadzący: dr inż. Sebastian Kozłowski

12 marca 2019
wer. 1.0

Spis treści

Streszczenie

Dokument ten zawiera szczegółowy opis zadania projektowego, który ma potwierdzić zrozumienie tematu przez autorów. Tematem zadania jest implementacja szybkiego i przenośnego generatora sieci bezskalowych w wybranym języku programowania. W sekcji (1) znajduje się omówienie tematu zadania, a w sekcjach (2) i (3) ogólny opis grafów losowych oraz opis modelu Barabasiego-Albert.

Sprawozdanie 1
Opis zadania projektowego

1 Opis zadania

Postawionym przed nami zadaniem jest zaimplementowanie generatora grafów losowych w ujęciu Barabasieg-Albert, które zostanie szczegółowo opisane w dalszej części dokumentu (3). Generator poza spełnianiem swojej podstawowej funkcji musi charakteryzować się przenośnością oraz jak najmniejszą złożnością obliczeniową i pamięciową. Dla poprawnie działającego generatora kolejnym krokiem w realizacji zadania jest zbadanie rozkładów stopni wierzchołków grafów i porównanie ich z modelami teoretycznymi. Pełna realizacja zadania zakłada istnienie możliwości zapisu wygenerowanego grafu do ustalonego formatu, co pozwoli odwtorzyć sam graf oraz przebieg eksperymentów numerycznych.
Dokumentacja ta jest wolna od opisu implementacji generatora, eksperymentów numerycznych oraz sposobu ich wizualizacji. Kwestie te zostaną omówione w kolejnych wersjach dokumentacjach, tj. odpowiednio: wersji 2.0 oraz 3.0.

2 Grafy losowe

Stosowanie teorii grafów do modelowania zjawisk zachodzących w realnym świecie jest oparte w dużej mierze na grafach losowych. Podyktowane jest to faktem, że zjawiska te i towarzyszące im zdarzenia wykazują w skali makroskopowej charakter stochastyczny. Przykłady dziedzin, w których stosowane są grafy losowe do modelowania pewnych zjawisk są następujące:

- sieci połączeń handlowych
- sieci `WWW`
- sieci neuronowe (rekurencyjne)
- sieci społecznościowe (np. `Facebook`)

Zdefiniowanie grafu losowego wymaga z kolei zdefiniowania struktur jak *przestrzeń grafów losowych* $\mathcal{G}$, która jest wyposażona w unormowaną miarę $\mathbb{P}(\bullet)$ mówiącą o prawdopodobieństwie wylosowania grafu G o pewnych właściwościach [3].
Zadanie to ze względu na złożoną strukturę obiektów jakimi są grafy nie jest tak intuicyjne jak określenie przestrzeni probabilistycznej dla zdarzeń, które można reprezentować liczbami. Z tego względu istnieje szereg alternatywnych modeli, które podejmują się rozwiązania tego zadania. Pokrótce zostanie omówiony najstarszy i najprostszy model wprowadzony przez Erdős'a i Rényi'e jeszcze w latach 60. ubiegłego wieku [2].

2.1 Model E-R

Model ten jest oparty o dwójkę parametrów (n, p): parametr $n \in \mathbb{N}$ oznacza liczbę wierzchołków generowanego grafu G, a $p \in (0, 1)$ stanowi o prawdopodobieństwie zdarzenia polegającego na zaistnieniu krawędz między każdą parą z n^2 wierzchołków grafu G.
Na podstawie powyższego łatwo widać, że rozkład stopni wierzchołków w grafie zadany jest przez rozkład dwumianowy z funkcją gęstości prawdopodobieństwa

$$p(n, k; p) = \binom{n-1}{k} p^k (1-p)^{n-k-1} \tag{1}$$

implikuje to fakt, że średni stopień wrzechołka $\mathbb{E}deg(v)$ wynosi $(n-1)p$. Ponadto, prawdopodobieństwo wylosowania grafu E-R o e krawędziach i n wierzchołkach wynosi $\binom{\binom{n}{2}}{m}p^m(1-p)^{\binom{n}{2}-m}$. Na tej podstawie liczba wszystkich możliwych grafów E-R o n wierzchołkach wynosi

$$\sum_{e=0}^{\binom{n}{2}} \binom{\binom{n}{2}}{e} = 2^{\binom{n}{2}} \tag{2}$$

przy czym $\binom{\binom{n}{2}}{e}$ oznacza liczbę e-elementowych kombinacji zbioru utworzonego ze wszystkich par zbioru n-elementowego.
Niestety, model taki nie jest najlepszym kandydatem do *naśladowania* obiektów rzeczywistych. Przy $p << 1$ rozkład stopni wierzchołków dany jest rozkładem Poissona, tj. rozkładem, który stosowany jest do zdarzeń rzadkich występujących w określonym przedziale czasu. Grafy generowane w tym modelu nie są w stanie dobrze odwzorowywać *hub*-ów, tj. skupisk.
Modele, które są wolne powyżej opisanych wad grafów opartych o model E-R, oparte są o rozkłady potęgowe i zostaną opisane w następnej sekcji (3).

3 Model Barabasiego-Albert

Jak się okazało, do modeli sieci rzeczywistych, np. sieci WWW czy sieci społecznościowych nie bardzo pasuje model E-R. W kontekście sieci WWW(gdzie *połączenie* między węzłami A i B jest zdefiniowane jako umieszczenie hipertextowego linka na stronie A do strony B) wynika to z tego, że autorzy tych stron z większym prawdopodobieństwem umieszczają linki na strony bardziej popularne, niż na te mniej popularne. Czyli, strony popularne stają się jeszcze bardziej popularniejsze, czyli "rich getting richer and poor poorer". Podobne rozumowanie odpowiada siecią społecznościowym czy połączenią handlowym. Odpowiada to tzw. sieciom bezskalowym. [1]
TODO: nawiązanie do poprzedniego rozdziału i napisanie motywacji modeli BA w kontekście zasady maksymalnej e

3.1 Sieć bezskalowa

Siecią bezskalową nazywamy sieć, rozkład liczby połączeń między węzłami jest wykładniczy, czyli spełnia równanie

$$P(k) \sim k^{-\gamma} \tag{3}$$

gdzie γ nazywana jest wartością właściwą grafu i zwykle mieści się w przedziale $(2,3)$ Oznacza to, że w sieci będziemy mieli dużo wierzchołków o małym stopniu i małą(w proporcji) liczbę wierzchołków o dużym stopniu.

3.2 Preferencyjne dołączanie wierzchołków

Mechanizm preferencyjnego dołączania wierzchołków polega na tym, że nowy wierzchołek z większym prawdopodobieństwem zostanie dołączony do starszego wierzchołka z większym stopniem:

$$P(k_i) = \frac{k_i}{\sum_{j=1}^{n} k_j} \tag{4}$$

Jest to tak zwana liniowa reguła preferencyjnego dołączania. W kontekście generatora sieci BA, w kolejnych krokach czasowych $t = 1, 2, 3$ przy dołączaniu stałej ilości wierzchołków m, prawdopodobieństwo będzie wyglądało następująco:

$$P(k_i) = \frac{k_i}{\sum_{i=1}^{t} k_i} = \frac{k_i}{2mt} \tag{5}$$

gdzie m - parametr sieci, *TODO: opisanie na czym polega ten mechanizm*

3.3 Rozkład stopni wierzchołków

Celem pracy jest napisanie generatora grafów modelu BA. Jednak napisany generator musi być zweryfikowany, czyli musimy stwierdzić że wygenerowane sieci rzeczywiście należą do modelu BA. Zrobimy to na podstawie weryfikacji rozkładu stopni wierzchołków grafów. Żeby to zrobić jednak musimy wyznaczyć teoretyczny rozkład wierzchołków. Można stosować do tego dwie metody: *TODO: wyprowadzenie zależności na rozkład stopni wierzchołków i napisanie, że są różne*

3.3.1 Model czasu ciągłego

Metoda czasu ciągłego polega na założeniu ciągłości czasu oraz na przyjęciu że stopnie wierzchołków $k_i(t)$ zmieniają się w czasie w sposób ciągły na ułamek stopnia(dlatego w tej metodzie posługujemy się stopniami/rozkładami uśrednionymi).

Zatem w pojedynczym kroku czasowym, stopień wierzchołka k zmienia się o

$$\Delta k_i/\Delta t \simeq \partial k_i/\partial t \tag{6}$$

TODO: wyprowadzenie r.s.w. matematyczne dla modelu czasu ciągłego Wynika z tego, że rozkład stopni wierzchołków ma następującą postać

$$P(k) = \frac{2m^2}{k^3} \tag{7}$$

3.3.2 Model równania *master*

Przy wykorzystaniu równania *master* wyznaczamy ścisły rozkład prawdopodobieństwa wierzchołków, dlatego że operujemy na stopniach węzłów a nie średnich stopni węzłów.
W tej metodzie posługujemy się równaniem mistrza, które opisuję zmianę w czasie ciągłym stanu układu fizycznego(w naszym przypadku - grafa). Chcemy uzyskać zatem równanie różniczkowe pierwszego rzędu, opisujące zmianę w czasie prawdopodobieństwa P_i znalezienia układu w stanie i, przy tym że w każdej chwili czasowej układ może zmienić swój stan i na dowolny j: $i \to j$. Zakładamy, że tempo takich zmian jest $T_{i\to j}$:

$$\frac{dP_i}{dt} = \sum_j P_j T_{j\to i} - \sum_j P_i T_{i\to j} \tag{8}$$

TODO: add math Rozwiązując równanie rekurencyjne dostajemy rozkład wierzchołków w sieci BA:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)} \tag{9}$$

przy czym $k \geqslant m$ Zauważmy różnicę rozkładów, którą jednak da się wytłumaczyć tym że równanie master jest metodą ścisłą, gdzie przybliżenia wynikały z poszukiwaniem granicznego rozkładu prawdopodobieństwa $P(k)$ dla $t \to \infty$

Sprawozdanie 2
Implementacja oraz eksperymenty numeryczne

Sprawozdanie 3
Wnioski końcowe

Literatura

[1] Model barabasiego-albert (ba). http://www.if.pw.edu.pl/~agatka/moodle/modele_SE_BA.html.

[2] P. Erdős and A Rényi. On the evolution of random graphs. In *PUBLICATION OF THE MATHEMATICAL INSTITUTE OF THE HUNGARIAN ACADEMY OF SCIENCES*, pages 17–61, 1960.

[3] Agata Fronczak. Wykładnicze grafy przypadkowe: teoria, przykłady, symulacje numeryczne. 2014.