
Wykładnicze grafy przypadkowe: teoria, przykłady, symulacje numeryczne

Agata Fronczak

Streszczenie

Omówione w tej pracy, podejście do modelowania sieci złożonych wykorzystujące wykładnicze grafy przypadkowe umożliwia tworzenie sieci o dowolnych (z góry ustalonych) własnościach strukturalnych. W grafach tych, wspomniane własności strukturalne są w prosty sposób określone przez podanie tzw. hamiltonianu sieci. W tej pracy przedstawiono teoretyczne podstawy modelu, oraz trzy podstawowe przykłady zastosowania wykładniczych grafów przypadkowych: i. do badania klasycznych grafów przypadkowych, ii. w analizie sieci o zadanym rozkładzie stopni węzłów, oraz iii. do tworzenia modeli sieci o strukturze blokowej. Z myślą o tych czytelnikach, którzy podejście wykładniczych grafów przypadkowych chcieliby wykorzystać w swoich badaniach naukowych, w pracy umieszczono obszerny rozdział poświęcony symulacjom numerycznym Monte Carlo, w którym omówiono algorytm Metropolis i jego wykorzystanie do badania wykładniczych grafów przypadkowych.

1.1. Wprowadzenie - modele sieci złożonych

W ostatniej dekadzie badania sieci złożonych stały się jednym z najciekawszych przykładów interdyscyplinarnych zastosowań matematyki dyskretnej i fizyki statystycznej [1–4]. Zbudowane z setek, tysiący, a nawet milionów elementów, pełniących różnorakie funkcje, powiązanych w skomplikowany, a jednak precyzyjny sposób, sieci¹,

¹W tej pracy, pojęcia graf i sieć są używane zamiennie, tak samo, jak w większości książek i prac naukowych poświęconych sieciom złożonym.

które rosną, dopasowują się do zmian otoczenia, optymalizują swoje działanie, opanowały wyobraźnię naukowców z wielu dziedzin. Jedną z przyczyn, dla których badania nad sieciami rzeczywistymi (komputerowymi, genetycznymi, społecznymi, ekonomicznymi etc.) stały się tak atrakcyjne, było odkrycie tego, że mimo funkcjonalnej różnorodności wymienione układy mają wiele wspólnych cech. I chociaż spośród tych cech najbardziej spektakularny był odkryty przez A.-L. Barabásiego i R. Alberta, opisany w magazynie Science [5], wspólny wielu sieciom, potęgowy (bezskałowy) rozkład stopni wierzchołków, dzisiaj wiadomo, że prawa potęgowe pojawiają się w nauce o sieciach złożonych niemal na każdym kroku ².

Wszechobecność praw potęgowych w sieciach rzeczywistych, jak również ich znaczenie dla tych układów starano się zrozumieć tworząc teoretyczne modele sieci. W wielu przypadkach, posługując się uproszczonymi modelami, pokazano, że struktura sieci i jej funkcjonalność są ze sobą ściśle powiązane. W ten sposób, modele, obok analizy sieci rzeczywistych, stały się jednym z głównych filarów, na których opiera się nauka o sieciach złożonych.

Oczywiście, istnieje wiele różnych modeli sieci złożonych. Najbardziej naturalną klasyfikacją tych modeli jest ich podział na sieci deterministyczne i przypadkowe ³. Do sieci przypadkowych zaliczamy te wszystkie sieci rzeczywiste i ich modele teoretyczne, których ewolucja lub procedura konstrukcyjna dopuszcza pewien czynnik losowości. Z formalnego punktu widzenia oznacza to, że mówiąc o jakimś modelu sieci przypadkowej, nie powinniśmy mieć na myśli żadnej konkretnej sieci, tylko zbiór różnych sieci, którego elementy mają różne prawdopodobieństwa realizacji. Ta cecha sieci przypadkowych, w sposób istotny odróżnia je od sieci deterministycznych, których procedury konstrukcyjne nie dopuszczają żadnej losowości.

W tej pracy omówione zostaną tzw. wykładnicze grafy przypadkowe ⁴. Rozważania teoretyczne na temat wykładniczych grafów przypadkowych zostaną zilustrowane przykładami sieci o różnych własnościach strukturalnych. W szczególności, omówione zostaną sieci o zadanej średniej liczbie krawędzi, które są równoważne klasycznym

²Oprócz wspomnianych rozkładów stopni węzłów, prawa potęgowe opisują również takie cechy sieci rzeczywistych jak: rozkłady wag połączeń oraz sił węzłów [6], zależność lokalnego współczynnika gronowania od stopnia węzła [7], a nawet samopodobny [9] i blokowy charakter sieci [8].

³Przykładem rzeczywistej sieci deterministycznej jest przypominający regularną siatkę układ ulic na nowojorskim Manhattanie. Nazwy ulic: ... East 46th, East 47th, ..., dobitnie wskazują, że ktoś to wszystko zaplanował. Zupełnie inny charakter ma układ ulic gdańskiej starówki. Uliczki mają różną długość, przecinają się pod różnymi kątami, sugerując, że układ ulic powstawał w sposób mniej lub bardziej przypadkowy.

⁴Obszerne opracowanie na temat wykładniczych grafów przypadkowych można znaleźć w książce autorstwa M.E.J. Newman [1] oraz w pracach [10, 11]. Uwadze czytelnika, który chciałby zapoznać się szerzej z omawianą tematyką polecamy również prace P.W. Hollanda i S. Leinhardta [12] oraz J.E. Besaga [13], w których idea wykładniczych grafów przypadkowych pojawiła się po raz pierwszy, a także prace O. Franka i D. Straussa [14, 15], które w znaczący sposób przyczyniły się do rozwinięcia i upowszechnienia formalizmu wykładniczych grafów przypadkowych.

grafom przypadkowym P. Erdösa i A. Rényi [18, 19], sieci o ustalonej sekwencji stopni węzłów [20] oraz sieci o strukturze modułowej [21]. Na koniec, omówiona zostanie metoda Monte Carlo umożliwiająca symulacje numeryczne wykładniczych grafów przypadkowych.

1.2. Wykładnicze grafy przypadkowe

1.2.1. Definicja modelu - zasada maksymalnej entropii

W badaniach sieci rzeczywistych dużym problemem jest to, że (zazwyczaj) istnieje tylko jedna realizacja (kopia) interesującej nas sieci. I tak, jest tylko jeden Internet, jedna sieć WWW, jedna sieć międzynarodowej wymiany handlowej. Dysponując danymi dotyczącymi struktury połączeń w takich pojedynczych sieciach (np. znając ich macierze sąsiedztwa) możemy oczywiście wyznaczyć podstawowe charakterystyki strukturalne. Możemy, na przykład, obliczyć średni stopień węzła, średnią odległość między dowolną parą węzłów, współczynnik gronowania sieci⁵ itd. Możemy nawet wyznaczyć rozkład stopni węzłów w tych sieciach. Dysponując informacjami nt. pojedynczej sieci nie możemy jednak odgadnąć reguł rządzących jej ewolucją, by później wykorzystać te reguły do stworzenia modelu sieci. Nie możemy również stwierdzić, które cechy są, potocznie mówiąc, ważniejsze od innych, które cechy determinują pozostałe. W takich właśnie sytuacjach doskonale sprawdza się model wykładniczych grafów przypadkowych, który pozwala utworzyć sieci o dowolnych (tj. z góry ustalonych) cechach strukturalnych bez wnikania w mechanizmy warunkujące powstawanie sieci rzeczywistych o podobnych własnościach. Ponadto, dysponując sieciami o zadanych cechach, można zbadać, jak te cechy wpływają na inne (strukturalne lub dynamiczne) własności badanych układów.

W modelu wykładniczych grafów przypadkowych wykorzystuje się idee zespołów statystycznych. W odniesieniu do sieci przypadkowych, pojęcie zespołu statystycznego oznacza zbiór sieci, $\mathcal{G} = \{G\}$ (np. zbiór wszystkich grafów prostych o ustalonej liczbie wierzchołków, N), w którym każda sieć ma określone prawdopodobieństwo realizacji, $P(G)$. W wykładniczych grafach przypadkowych prawdopodobieństwo to jest opisane funkcją wykładniczą⁶,

$$P(G) \propto e^{H(G)}, \quad (1.1)$$

⁵Współczynnik gronowania, C_i , pewnego węzła i wyraża prawdopodobieństwo, że najbliżsi sąsiedzi tego węzła względem siebie są również najbliższymi sąsiadami. Z tego opisu wynika, że współczynnik gronowania jest równy: $C_i = E_i / \binom{k_i}{2}$, gdzie E_i jest liczbą połączeń między najbliższymi sąsiadami węzła i , k_i jest stopniem tego węzła, zaś $\binom{k_i}{2}$ opisuje liczbę wszystkich połączeń, które mogłyby istnieć między jego sąsiadami (zob. rozdz. 2. *Własności sieci rzeczywistych* w [3]).

⁶Stąd też pochodzi nazwa modelu: wykładnicze grafy przypadkowe (ang. exponential random graphs).

gdzie $H(G)$ jest tak zwanym strukturalnym hamiltonianem zespołu. Wykładnicza postać rozkładu prawdopodobieństwa $P(G)$ nie jest przypadkowa. W dalszej części tego podrozdziału pokażemy, że taka postać prawdopodobieństwa maksymalizuje entropię [22, 23] zespołu sieci przypadkowych,

$$S(P(G)) = - \sum_{G \in \mathcal{G}} P(G) \ln P(G), \quad (1.2)$$

przy określonych więzach, reprezentujących, na przykład, określone cechy strukturalne sieci, które to więzy znajdują odzwierciedlenie w hamiltonianie zespołu. Innymi słowy, wykładnicze grafy przypadkowe to najbardziej losowe spośród sieci o zadanych własnościach.

Niech zatem \mathcal{G} oznacza zbiór możliwych realizacji badanych sieci przypadkowych, zaś $P(G)$ niech będzie prawdopodobieństwem pewnego grafu z tego zbioru, $G \in \mathcal{G}$. Chcielibyśmy wybrać $P(G)$ w taki sposób, by wartości oczekiwane pewnych, wybranych a priori, strukturalnych parametrów sieci ⁷, np.

$$\{x_i(G)\} = x_1(G), x_2(G), \dots, x_r(G), \quad (1.3)$$

gdzie r jest liczbą tych parametrów, miały określone wartości, tzn. by były odpowiednio równe:

$$\{x_i^*\} = x_1^*, x_2^*, \dots, x_r^*. \quad (1.4)$$

Powyższe warunki (więzy) można zapisać w postaci równań:

$$\forall_{i=1,2,\dots,r} \quad \langle x_i \rangle = \sum_{G \in \mathcal{G}} x_i(G) P(G) = x_i^*. \quad (1.5)$$

Zgodnie z zasadą maksymalnej entropii [22, 23], szukany rozkład $P(G)$ powinien maksymalizować wyrażenie (1.2), przy warunkach (1.5) oraz przy warunku unormowania rozkładu. W celu znalezienia takiego rozkładu można posłużyć się metodą mnożników Lagrange'a. Wprowadzając mnożniki, $\{\theta_i\} = \theta_1, \theta_2, \dots, \theta_r$ odpowiadające narzuconym na rozkład więzom oraz mnożnik α odpowiadający warunkowi unormowania rozkładu, można pokazać, że maksymalną wartość entropii uzyskuje się dla rozkładu spełniającego poniższą zależność:

$$\frac{\partial}{\partial P(G)} \left[S - \alpha \left(1 - \sum_{G \in \mathcal{G}} P(G) \right) - \sum_{i=1}^r \theta_i \left(x_i^* - \sum_{G \in \mathcal{G}} x_i(G) P(G) \right) \right] = 0, \quad (1.6)$$

⁷Takimi parametrami mogą być, na przykład, liczba krawędzi w sieci, $E(G)$, średnia droga między dowolną parą węzłów, $l(G)$, liczba trójkątów (tj. cykli o długości 3), $T(G)$, itd.

którą można przedstawić w uproszczonej postaci,

$$-\ln P(G) - 1 + \alpha + \sum_{i=1}^r \theta_i x_i(G) = 0, \quad (1.7)$$

z której uzyskujemy wykładniczy rozkład prawdopodobieństwa (por. (1.1))

$$P(G) = \frac{e^{H(G)}}{Z}. \quad (1.8)$$

W wyrażeniu (1.8) funkcja

$$H(G) = \sum_{i=1}^r \theta_i x_i(G) \quad (1.9)$$

jest hamiltonianem zespołu spełniającego więzy postaci (1.5), zaś

$$Z = e^{1-\alpha} = \sum_{G \in \mathcal{G}} e^{H(G)} \quad (1.10)$$

nosi nazwę sumy statystycznej, którą można wyznaczyć z warunku unormowania rozkładu, $\sum_{G \in \mathcal{G}} P(G) = 1$.

1.2.2. Średnie po zespole i fluktuacje

Znając rozkład prawdopodobieństwa, $P(G)$, określony na zbiorze \mathcal{G} można wyznaczyć wartości oczekiwane różnych parametrów strukturalnych badanych sieci. Na przykład, w zespole wykładniczych grafów przypadkowych o hamiltonianie (1.9), wartość oczekiwana parametru y będzie wynosiła:

$$\langle y \rangle = \sum_{G \in \mathcal{G}} y(G) P(G) = \frac{1}{Z} \sum_{G \in \mathcal{G}} y(G) e^{H(G)}. \quad (1.11)$$

Wyrażenie to pozwala uzyskać najlepsze oszacowanie dla wartości parametru y w sieciach, w których pewne własności strukturalne są znane, tzn. takich sieciach, które spełniają równania więzów (1.5). W szczególności, korzystając z wyrażenia (1.11) możemy oszacować strukturalne własności sieci rzeczywistych, o których mamy niepełną informację, w postaci, np. znanego rozkładu stopni wierzchołków.

Korzystając z zależności (1.11) można również otrzymać wyrażenia opisujące, w jaki sposób oczekiwane wartości podstawowych parametrów definiujących badany zespół⁸ zależą od wartości sprzężonych z nimi mnożników Lagrange'a, tj.

$$\langle x_i \rangle = \frac{1}{Z} \sum_{G \in \mathcal{G}} x_i(G) e^{H(G)} = \frac{1}{Z} \frac{\partial}{\partial \theta_i} \sum_{G \in \mathcal{G}} e^{H(G)} = \frac{1}{Z} \frac{\partial Z}{\partial \theta_i} = \frac{\partial F}{\partial \theta_i}, \quad (1.12)$$

⁸porównaj równania więzów (1.5) oraz hamiltonian zespołu (1.9);

gdzie $H(G)$ jest opisane wyrażeniem (1.9), zaś

$$F = \ln Z, \quad (1.13)$$

będąca funkcją parametrów zespołu, $F = F(\theta_1, \dots, \theta_r)$, jest nazywana energią swobodną⁹.

Z wyrażenia (1.12) wynika, że średnia wartość parametru x_i , tzn. $\langle x_i \rangle$, jest równa pochodnej energii swobodnej, F , po mnożniku θ_i . W podobny sposób można pokazać, że druga pochodna F po θ_i charakteryzuje fluktuacje (rozrzut) parametru x_i wokół jego wartości średniej:

$$\langle x_j^2 \rangle - \langle x_j \rangle^2 = \frac{1}{Z} \frac{\partial^2 Z}{\partial \theta_j^2} - \frac{1}{Z^2} \left(\frac{\partial Z}{\partial \theta_j} \right)^2 = \frac{\partial}{\partial \theta_j} \left(\frac{1}{Z} \frac{\partial Z}{\partial \theta_j} \right) = \frac{\partial^2 F}{\partial \theta_j^2} = \frac{\partial \langle x_j \rangle}{\partial \theta_j}. \quad (1.14)$$

W fizyce statystycznej, powyższe wyrażenie jest znane pod nazwą relacji fluktuacyjno-dysypacyjnej lub jako twierdzenie fluktuacje-odpowiedź. Nazwa ta jest związana z tym, że lewa strona wyrażenia (1.14) opisuje fluktuacje parametru x_i , podczas gdy jej prawa strona charakteryzuje podatność tego parametru, tzn. opisuje, w jaki sposób jego średnia wartość, $\langle x_i \rangle$, zmieniałaby się, gdybyśmy zmienili sprzężony z nią mnożnik Lagrange'a, θ_i .

W tym miejscu warto zastanowić się, czym są mnożniki Lagrange'a i jaką rolę pełnią w sieciach rzeczywistych. Otóż na drugą część powyższego pytania, przynajmniej częściowo, odpowiedzieliśmy już wcześniej: Konkretnie mnożniki Lagrange'a odpowiadają konkretnym więzom narzuconym na badaną sieć. Na przykład, jeśli w sieci ważonej¹⁰ mnożnik θ_i jest związany z siłą¹¹, $s_i(G)$, wężła i oznacza to, że wartość tego mnożnika przekłada się (w sposób ściśle określony przez energię swobodną $F(\theta_1, \dots, \theta_r)$) na wartość $\langle s_i \rangle$, zob. Eq. (1.12). Odnosząc te rozważania to sieci rzeczywistych, tzn. poszukując mnożnika θ_i warunkującego siłę wężła, należałoby zastanowić się nad tym, jakie czynniki zewnętrzne mają wpływ na wagi krawędzi dołączonych do tego wężła. Na przykład, w artykułach [16, 17] nt. sieci handlu światowego, w której wężłami były poszczególne państwa, zaś krawędziom przypisano wartości rocznej bilateralnej wymiany handlowej, pokazano, że siły wężłów zależą jedynie od produktów krajowych brutto (PKB) poszczególnych państw.

⁹Termin energia swobodna został zapożyczony z fizyki statystycznej, podobnie zresztą jak termin suma statystyczna. Ogólnie, dla fizyków wykładnicze grafy przypadkowe są szczególnie proste (intuicyjnie zrozumiałe) ponieważ koncepcyjnie nawiązują one do zapoczątkowanej przez E.T. Jaynesa szkoły fizyki statystycznej opartej na zasadzie maksymalnej entropii [24–26], która jest formalnie równoważna tradycyjnej szkole fizyki statystycznej zapoczątkowanej przez Boltzmanna i Gibbsa.

¹⁰tj. takiej, w której każdej krawędzi przypisana jest waga;

¹¹W sieciach ważonych, siłą wężła nazywamy sumę wag krawędzi dołączonych do tego wężła: $s_i = \sum_{j=1}^N w_{ij}$.

Na koniec, wyjaśnijmy jeszcze, dlaczego ważne jest to, by znać zależność (analityczną bądź numeryczną) $\langle x_i \rangle$ od θ_i . Otóż, jak pokażemy w dalszej części tej pracy, znajomość takiej zależności umożliwia wykonywanie obliczeń analitycznych i/lub numerycznych z wykorzystaniem omawianych zespołów sieci.

1.3. Przykłady

1.3.1. Klasyczne grafy przypadkowe

Procedura konstrukcyjna i podstawowe własności

Klasyczne grafy przypadkowe (lub grafy ER) po raz pierwszy zostały omówione przez wybitnych matematyków węgierskiego pochodzenia: P. Erdösa i A. Rényiego [18, 19]. Grafy te, a ściślej ich statystyczną zbiorowość, tradycyjnie definiuje się podając procedurę konstrukcyjną, która obejmuje dwa etapy:

- i. Na początku ustala się liczbę węzłów N należących do grafu.
- ii. Następnie, z zadaniem prawdopodobieństwem p przypisuje się połączenia każdej z $\binom{N}{2}$ par węzłów.

Z procedury konstrukcyjnej klasycznych grafów przypadkowych wynika, że średnia liczba krawędzi w tych grafach wynosi

$$\langle E \rangle = p \binom{N}{2} = p \frac{N(N-1)}{2}, \quad (1.15)$$

skąd wnioskujemy, że średni stopień dowolnego węzła jest równy:

$$\langle k \rangle = \frac{2\langle E \rangle}{N} = p(N-1) \simeq pN. \quad (1.16)$$

Analizując procedurę konstrukcyjną grafów ER można również wywnioskować, że rozkład stopni wierzchołków charakteryzujący ten model sieci przypadkowych jest rozkładem dwumianowym¹², który w granicy $p \ll 1$ można przybliżyć rozkładem Poissona:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \simeq \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!}, \quad (1.17)$$

gdzie $\langle k \rangle$ (1.16) jest średnim stopniem węzła.

¹²Prawdopodobieństwo, że podczas konstrukcji grafu dowolny węzeł uzyska k na $N-1$ możliwych połączeń do innych wierzchołków sieci jest równe prawdopodobieństwu uzyskania k sukcesów w $N-1$ próbach, gdy prawdopodobieństwo sukcesu wynosi p , a prawdopodobieństwo porażki $1-p$.

Z uwagi na rozkład stopni węzłów (1.17) klasyczne grafy przypadkowe określa się niekiedy mianem grafów poissonowskich. To określenie pojawia się zazwyczaj w zestawieniu z określeniami: sieci potęgowe lub bezskalowe (czyli sieci o potęgowym rozkładzie stopni węzłów, $P(k) \propto k^{-\gamma}$), i służy uwypukleniu jednej z najważniejszych wad grafów ER. Wady, która uniemożliwia wykorzystanie tych sieci do modelowania rzeczywistych sieci złożonych. Wspomnianą wadą jest wąski, o bardzo małej wariancji, rozkład stopni wierzchołków. Przypomnijmy, że w rozkładzie Poissona wariancja σ_k^2 jest równa wartości średniej $\langle k \rangle$. Ta cecha rozkładu Poissona sprawia, że o ile w sieciach rzeczywistych węzły o stopniach znacznie różniących się od średniego stopnia $\langle k \rangle$ są często spotykane, w grafach ER takie węzły wcale się nie pojawiają. W tej sytuacji, naturalnym pytaniem jest: Dlaczego w licznych opracowaniach na temat sieci złożonych tak wiele miejsca poświęca się grafom ER. Powód jest prosty: Porównanie własności grafów ER z własnościami sieci bezskalowych pozwala zrozumieć, dlaczego sieci bezskalowe są wyjątkowe. W nawiązaniu zaś do tematu wykładniczych grafów przypadkowych prawdą jest to, że model ER pozwala w prosty sposób zilustrować ideę zespołów statystycznych sieci przypadkowych.

Zespół statystyczny klasycznych grafów przypadkowych

Na początek zauważmy, że postępując zgodnie z procedurą konstrukcyjną klasycznych grafów przypadkowych można utworzyć

$$2^{\binom{N}{2}} = \sum_{E=0}^{\binom{N}{2}} \binom{\binom{N}{2}}{E} \quad (1.18)$$

różnych realizacji takich grafów. Wyjaśnijmy, że umieszczony po prawej stronie wyrażenia (1.18) symbol Newtona $\binom{\binom{N}{2}}{E}$ odpowiada liczbie różnych realizacji grafów prostych o ustalonej liczbie krawędzi E , tj. liczbie wszystkich kombinacji E elementów zbioru $\binom{N}{2}$ elementowego. Sumując symbol Newtona po wszystkich wartościach parametru $E = 0, 1, 2, \dots, \binom{N}{2}$ otrzymujemy liczbę wszystkich możliwych grafów prostych o zadanej liczbie węzłów N i dowolnej liczbie krawędzi. Liczba ta odpowiada liczbie różnych, zero-jedynkowych, symetrycznych macierzy sąsiedztwa o rozmiarze $N \times N$. Nawet dla niewielkich rozmiarów sieci ta liczba jest ogromna. Na przykład, gdy $N = E = 100$, liczba grafów prostych o takich parametrach jest równa $\binom{\binom{100}{2}}{100} \approx 10^{211}$, zaś liczba wszystkich grafów prostych o $N = 100$ wierzchołkach wynosi $2^{\binom{100}{2}} \approx 10^{1490}$.

Z uwagi na to, że w klasycznych grafach przypadkowych prawdopodobieństwo pojawienia się dowolnej krawędzi jest równe p , różne realizacje grafów ze zbioru \mathcal{G} wszystkich grafów prostych o N wierzchołkach mają różne prawdopodobieństwa realizacji, $P(G)$. Na przykład, gdy $p = 1$, jedynym grafem $G \in \mathcal{G}$, dla którego $P(G) \neq 0$

jest graf pełny, w którym każdy węzeł jest połączony krawędzią ze wszystkimi pozostałymi węzłami grafu. Gdy jednak $p \neq \{0, 1\}$ prawdopodobieństwo realizacji grafu G zależy od liczby jego krawędzi $E(G)$ i jest równe:

$$P(G) = p^{E(G)}(1-p)^{\binom{N}{2}-E(G)}. \quad (1.19)$$

Wyrażenie to opisuje prawdopodobieństwo równoczesnej realizacji $\binom{N}{2}$ niezależnych zdarzeń, spośród których $E(G)$ zdarzeń (typu 'jest krawędź') ma prawdopodobieństwo realizacji równe p , zaś prawdopodobieństwo pozostałych zdarzeń (typu 'brak krawędzi') wynosi $1-p$. Oczywiście, średnia liczba krawędzi w badanym zespole grafów, tj.

$$\langle E \rangle = \sum_{E(G)=0}^{\binom{N}{2}} E(G)P(G) = p \binom{N}{2}, \quad (1.20)$$

wyraża się taką samą zależnością, co średnia liczba krawędzi obliczona wcześniej (1.15).

Hamiltonian zespołu i średnie po zespole

Aby pokazać, że grafy ER o N węzłach i prawdopodobieństwie połączenia p można analizować z wykorzystaniem formalizmu wykładniczych grafów przypadkowych, zauważmy, że prawdopodobieństwo grafu w tym zespole opisane wyrażeniem (1.19) można przepisać w wykładniczej postaci:

$$P(G) = \left(\frac{p}{1-p} \right)^{E(G)} (1-p)^{\binom{N}{2}} = \frac{e^{\theta E(G)}}{Z}, \quad (1.21)$$

gdzie

$$\theta = \ln \left[\frac{p}{1-p} \right] \quad \text{oraz} \quad p = \frac{e^{\theta}}{1+e^{\theta}}, \quad (1.22)$$

zaś

$$Z = \left(\frac{1}{1-p} \right)^{\binom{N}{2}} = (1+e^{\theta})^{\binom{N}{2}}. \quad (1.23)$$

Z wyrażenia (1.21) wynika, że zespół klasycznych grafów przypadkowych jest równoważny zespołowi wykładniczych grafów przypadkowych o hamiltonianie

$$H(G) = \theta E(G). \quad (1.24)$$

Powyższa postać hamiltonianu sugeruje, że grafy te można traktować jak zespół sieci losowych (ang. maximum entropy random networks) o ustalonej średniej liczbie krawędzi.

Przy okazji zauważmy, że chociaż suma statystyczna zespołu jest znana (1.23), można ją również wyznaczyć bezpośrednio z ogólnego wzoru (1.10). Aby to pokazać trzeba zauważyć, że liczbę krawędzi w dowolnej realizacji grafu $G \in \mathcal{G}$ można obliczyć sumując elementy macierzy sąsiedztwa $A_{ij}(G)$ badanego grafu, które, dla $A_{ij}(G) = 1$ (lub 0) reprezentują 'istnienie' (odpowiednio 'brak') krawędzi między węzłami i oraz j , tj.

$$E(G) = \sum_{i=1}^N \sum_{j=i+1}^N A_{ij}(G). \quad (1.25)$$

Podstawiając (1.25) do (1.24), a następnie korzystając ze wzoru (1.10) dostajemy:

$$Z = \sum_{G \in \mathcal{G}} e^{\theta E(G)} = \sum_{G \in \mathcal{G}} \prod_{i=1}^N \prod_{j=i+1}^N e^{\theta A_{ij}} = \prod_{i=1}^N \prod_{j=i+1}^N \sum_{A_{ij}=0}^1 e^{\theta A_{ij}} = (1 + e^{\theta})^{\binom{N}{2}}. \quad (1.26)$$

Podobnie, wykorzystując ogólną zależność (1.12) można pokazać, że średnia liczba krawędzi w badanym zespole sieci przypadkowych wynosi (porównaj z zależnościami (1.15) oraz (1.20)):

$$\langle E \rangle = \frac{\partial F}{\partial \theta} = \binom{N}{2} \frac{e^{\theta}}{1 + e^{\theta}} = \binom{N}{2} p, \quad (1.27)$$

gdzie energia swobodna, F , jest dana wzorem:

$$F = \ln Z = \binom{N}{2} \ln[1 + e^{\theta}]. \quad (1.28)$$

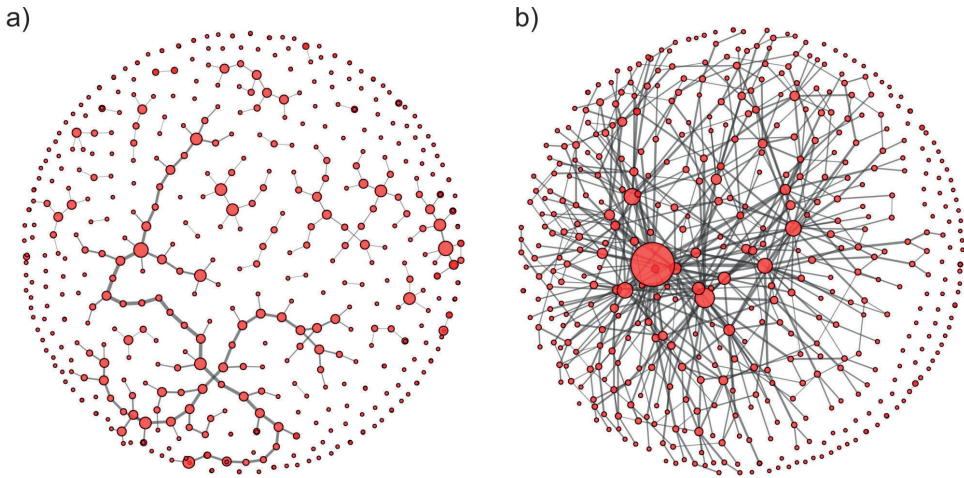
W rozdziale 1.4 pokażemy, w jaki sposób grafy ER można symulować numerycznie wykorzystując przy tym wyprowadzone w tym podrozdziale zależności matematyczne. Tymczasem, na rysunku 1.1a pokazujemy przykładową realizację grafu ER o liczbie węzłów $N = 480$ i średniej liczbie krawędzi $\langle E \rangle = 480$ (co odpowiada mnożnikowi $\theta = -6.17$).

1.3.2. Uogólnione grafy przypadkowe

Procedura konstrukcyjna i hamiltonian zespołu

Omawiając własności klasycznych grafów przypadkowych wspomnieliśmy, że poissonowski rozkład stopni węzłów (1.17) jest ich największą wadą jako potencjalnych modeli rzeczywistych sieci złożonych. Chcąc ominąć te ograniczenia zaproponowano model tzw. uogólnionych grafów przypadkowych¹³, znany również jako model kon-

¹³Czytelnika zainteresowanego teorią sieci złożonych z pewnością zainteresuje to, że większość ważnych wyników dotyczących sieci złożonych, związanych np. z modelowaniem epidemii w sieciach bezskalnych [28], badaniem odporności tych sieci na przypadkowe błędy węzłów [29] i ich podatności na celowe ataki [30], została uzyskana w oparciu o ten właśnie model, model grafów przypadkowych o zadanym rozkładzie stopni węzłów.



Rysunek 1.1: Przykładowe realizacje grafów wykładniczych a) grafu ER o liczbie węzłów $N = 480$ i średniej liczbie krawędzi $\langle E \rangle = 480$, b) uogólnionego grafu wykładniczego o liczbie węzłów $N = 480$ i o potęgowym rozkładzie oczekiwanych stopni węzłów $P(\langle k_i \rangle) \sim \langle k_i \rangle^{-3}$.

figuracyjny lub model grafów przypadkowych o ustalonym a priori rozkładzie stopni węzłów [27].

Procedura konstrukcyjna tych grafów jest następująca:

- i. Na początku, podobnie jak w grafach ER, ustala się liczbę węzłów N .
- ii. Następnie, do każdego węzła $i = 1, 2, \dots, N$ dołącza się k_i^* połówek krawędzi, przy czym liczby k_i^* są losowane z ustalonego a priori rozkładu $P(k)$.
- iii. W ostatnim kroku procedury konstrukcyjnej w sposób losowy łączy się ze sobą połówki krawędzi.

W wyniku losowego łączenia krawędzi, realizowanego w ostatnim etapie procedury konstrukcyjnej, można uzyskać wiele różnych realizacji sieci o zadanej sekwencji stopni węzłów ¹⁴:

$$\{k_1^*, k_2^*, \dots, k_N^*\}. \quad (1.29)$$

Z powyższej dyskusji wynika zatem, że badając model uogólnionych grafów przypadkowych, w rzeczywistości mamy do czynienia z zespołem sieci przypadkowych.

¹⁴Sekwencja (1.29) to zbiór liczb losowych z rozkładu $P(k)$. Taką sekwencję należy interpretować w następujący sposób: węzeł o numerze 1 ma stopień równy k_1^* , węzeł o numerze 2 ma stopień równy k_2^* , itd.

Co więcej, zespół ten powinien być formalnie równoważny zespołowi wykładniczych grafów przypadkowych o hamiltonianie postaci

$$H(G) = \sum_{i=1}^N \theta_i k_i(G), \quad (1.30)$$

w którym więzami są ustalone stopnie kolejnych węzłów, tj.

$$\forall_{i=1,2,\dots,N} \quad \langle k_i \rangle = \sum_{G \in \mathcal{G}} k_i(G) P(G) = k_i^*. \quad (1.31)$$

Suma statystyczna i średnie po zespole

Okazuje się, że sumę statystyczną rozważanego zespołu można, podobnie jak w grafach ER, wyznaczyć w sposób ścisły. Aby to zrobić, hamiltonian zespołu należy wyrazić poprzez elementy macierzy sąsiedztwa. W tym celu zauważmy, że stopień k_i dowolnego węzła i można zapisać w postaci:

$$k_i(G) = \sum_{j=1}^N A_{ij}(G). \quad (1.32)$$

Podstawiając powyższą zależność do hamiltonianu zespołu (1.30), a potem nową postać hamiltonianu wstawiając do ogólnego wzoru na sumę statystyczną (1.10) dostajemy (zob. równ. (1.26)):

$$Z = \prod_{i=1}^N \prod_{j=i+1}^N (1 + e^{(\theta_i + \theta_j)}), \quad (1.33)$$

oraz

$$F = \sum_{i=1}^N \sum_{j=i+1}^N \ln [1 + e^{(\theta_i + \theta_j)}]. \quad (1.34)$$

Następnie, różniczkując energię swobodną F względem θ_i otrzymujemy wyrażenie opisujące, w jaki sposób średni stopień i -tego węzła zależy od parametrów zespołu, tj.

$$\langle k_i \rangle = \frac{\partial F}{\partial \theta_i} = \sum_{j=1}^N \frac{1}{1 + e^{(\theta_i + \theta_j)}} = \sum_{j=1}^N p_{ij}, \quad (1.35)$$

gdzie

$$p_{ij} = \langle A_{ij} \rangle = \frac{1}{1 + e^{(\theta_i + \theta_j)}} \quad (1.36)$$

jest prawdopodobieństwem istnienia krawędzi między węzłami i oraz j .

Z powyższych rozważań wynika, że przepisując zależność (1.35) dla wszystkich węzłów w sieci $i = 1, 2, \dots, N$ otrzymamy układ N równań z N niewiadomymi: $\theta_1, \theta_2, \dots, \theta_N$. Oznacza to, że jeśli znamy oczekiwane stopnie wszystkich węzłów (1.31), to podstawiając je do wspomnianych równań i rozwiązując te równania (np. numerycznie) uzyskamy wartości wszystkich parametrów $\{\theta_i\}$ charakteryzujących badany zespół. Dlaczego mielibyśmy zwracać sobie głowę wyznaczeniem mnożników Lagrange'a? Otóż dlatego, ponieważ znając te parametry moglibyśmy, na przykład, stworzyć numeryczny model sieci¹⁵, a następnie na uzyskanej strukturze sieciowej moglibyśmy zaimplementować dowolny proces dynamiczny, reprezentujący, na przykład, rozprzestrzenianie się epidemii lub błędzenie przypadkowe po sieci. Ponadto, znając wartości $\{\theta_i\}$ odpowiadające badanym sieciom moglibyśmy wykonać obliczenia (analityczne lub numeryczne) mające na celu, na przykład, oszacowanie krytycznego tempa rozprzestrzeniania się epidemii lub średniego czasu dotarcia pakietu do celu w sieci komputerowej.

Przybliżenie rzadkich sieci

Okazuje się, że w granicy rzadkich sieci wyrażenia (1.35) i (1.36) można znacząco uprościć. Z założenia o 'rzadkich sieciach' wynika bowiem, że w takich sieciach prawdopodobieństwo połączenia jest bardzo małe, $p_{ij} \ll 1$, co oznacza, że w mianowniku wyrażenia (1.36) można pominąć 1 w porównaniu z drugim składnikiem sumy. Dzięki temu, w granicy rzadkich sieci, prawdopodobieństwo istnienia połączenia między węzłami i oraz j faktoryzuje się:

$$p_{ij} \simeq e^{-\theta_i} e^{-\theta_j}. \quad (1.37)$$

Następnie, podstawiając (1.37) do wyrażenia (1.35) otrzymujemy prostą zależność między średnim stopniem węzła i oraz przypisanym mu mnożnikiem Lagrange'a:

$$\langle k_i \rangle \simeq e^{-\theta_i} \sum_{j=1}^N e^{-\theta_j} = e^{-\theta_i} \sqrt{\langle k \rangle N}, \quad (1.38)$$

gdzie wykorzystaliśmy lemat o uściskach dłoni, $\sum_i \langle k_i \rangle = \langle k \rangle N$. Rysunek 1.1b przedstawia przykładową realizację uogólnionego grafu przypadkowego o potęgowym rozkładzie oczekiwanych stopni węzłów. Na koniec zauważmy, że podstawiając zależność (1.38) do (1.37) dostajemy znane wyrażenie opisujące prawdopodobieństwo istnienia krawędzi w uogólnionych grafach przypadkowych:

$$p_{ij} = \frac{\langle k_i \rangle \langle k_j \rangle}{\langle k \rangle N}. \quad (1.39)$$

¹⁵Zobacz rozdz. 1.4 poświęcony symulacjom numerycznym metodą Monte Carlo.

1.3.3. Modele sieci blokowych

Sieci rzeczywiste o strukturze blokowej

Wiele sieci rzeczywistych ma strukturę blokową (modułową) [31, 35]. Oznacza to, że w strukturze sieci można wyróżnić grupy węzłów takie, że gęstość połączeń między węzłami należącymi do tej samej grupy (modułu) jest znacznie większa niż gęstość połączeń między węzłami należącymi do różnych grup (modułów). Przykładami modułów w rzeczywistych sieciach złożonych są: grupy zaprzyjaźnionych ze sobą osób w sieciach społecznych, kraje uczestniczące w międzynarodowej wymianie handlowej, które podpisały regionalne porozumienia handlowe, grupy stron internetowych zawierających treści o tej samej tematyce, czy też białka, które w sieci protein współtworzą określone szlaki metaboliczne. Powyższe przykłady pokazują, że w wielu sieciach rzeczywistych przynależność węzłów do bloków może być (i zwykle jest) związana z dobrze określonymi własnościami tych węzłów, które mogą mieć wpływ na funkcjonowanie całej sieci, a nie tylko określonych grup węzłów.

Z tego powodu, struktura blokowa jest uważana za jedną z najważniejszych, obok bezskalowych rozkładów stopni węzłów i efektu małych światów¹⁶ cechę rzeczywistych sieci złożonych. Z tego również powodu, w ostatnich latach różne zagadnienia związane z tą cechą sieci były intensywnie badane, chociaż z perspektywy czasu wiadać, że badania te, jak dotąd, skupiały się przede wszystkim na opracowaniu efektywnych algorytmów do detekcji modułów w sieciach rzeczywistych [36]. W tym kontekście, w charakterze środowiska do testowania wspomnianych algorytmów, powszechnie były wykorzystywane tzw. klasyczne modele blokowe [37], które ostatnio zaczęto zastępować modelami blokowymi o potęgowych rozkładach stopni węzłów [8, 38].

Klasyczny model blokowy i model blokowy z korektą na stopnie węzłów

W klasycznym modelu blokowym [39], każdy spośród N węzłów sieci jest przypisany do jednego z K bloków. Bloki mają charakter grafów ER. O połączeniach między węzłami należącymi do tego samego bloku mówimy, że są to połączenia lokalne, odróżniając je w ten sposób od rzadszych połączeń globalnych, które łączą węzły należące do różnych modułów.

Wykładnicze grafy przypadkowe zostały po raz pierwszy wykorzystane do modelowania sieci o strukturze modułowej w pracy [21], w której oprócz klasycznego

¹⁶O efekcie *małych światów* [32] w sieciach złożonych mówimy wtedy, gdy średnia odległość, $\langle l \rangle$, między dowolną parą węzłów w tej sieci zmienia się z rozmiarem sieci, N , wolniej niż potęgowo. W klasycznych i uogólnionych grafach przypadkowych efekt małych światów objawia się tym, że $\langle l \rangle \sim \ln N$ [33]. W sieciach bezskalowych, w których wykładnik charakterystyczny γ w rozkładzie stopni węzłów $P(k) \sim k^{-\gamma}$ jest z przedziału $(2, 3)$, mamy do czynienia z tzw. efektem *ultra-małych światów*, tj. $\langle l \rangle \sim \ln \ln N$ [34].

modelu blokowego o hamiltonianie postaci:

$$H(G) = \sum_{r \leq s} \omega_{rs} E_{rs}(G), \quad (1.40)$$

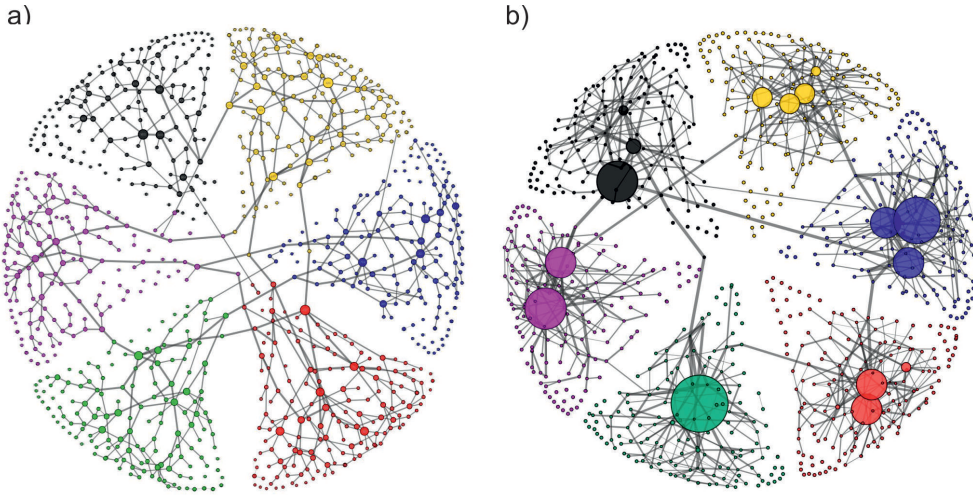
gdzie E_{rs} jest liczbą krawędzi między modułami r oraz s , zaś E_{rr} reprezentuje liczbę krawędzi wewnątrz modułu r , zbadano również model blokowy z tzw. korektą na stopnie węzłów o hamiltonianie postaci:

$$H(G) = \sum_i v_i k_i(G) + \sum_{r \leq s} \omega_{rs} E_{rs}(G), \quad (1.41)$$

gdzie k_i oznacza stopień wężła i . We wspomnianej pracy obliczono sumy statystyczne tych zespołów, dzięki czemu uzyskano wyrażenia opisujące, w jaki sposób wartości średnie $\langle E_{rs} \rangle$, $\langle E_{rr} \rangle$ oraz $\langle k_i \rangle$ charakteryzujące strukturalne własności badanych sieci zależą od parametrów: ω_{rs} , ω_{rr} oraz v_i . Oczywiście, mnożniki Lagrange'a v_i dla $i = 1, \dots, N$ warunkują odpowiednie stopnie węzłów badanej sieci (por. równ. (1.30)), zaś mnożniki ω_{rs} kontrolują liczbę krawędzi między modułami r i s .

W analizie modelu blokowego z korektą na stopnie węzłów, ważną ideą okazało się zdefiniowanie tzw. wewnętrznych i zewnętrznych stopni węzłów. Stopniem wewnętrznym, k_i^{int} , wężła i nazwano liczbę jego najbliższych sąsiadów należących do tego samego modułu, zaś stopniem zewnętrznym, k_i^{ext} , liczbę jego pozostałych sąsiadów, równą $k_i - k_i^{int}$. Pokazano, że w tym zespole sieci istnieje liniowa zależność między $\langle k_i^{int} \rangle$ oraz $\langle k_i^{ext} \rangle$, która jest koncepcyjnie podobna do relacji skalowania, które zostały zaobserwowane w samopodobnych (fraktalnych) rzeczywistych sieciach złożonych [9]. Pojawienie się tej zależności w wykładniczych grafach przypadkowych o dość prostym hamiltonianie (1.41) jest o tyle zaskakujące, że w pracach nt. samopodobnych sieci złożonych [9, 40] podobne zależności zostały zidentyfikowane, jako potencjalne źródło fraktalności wynikające z ewoluującego charakteru badanych sieci.

Na rysunku 1.2 przedstawiono przykładowe realizacje sieci należących do omawianych zespołów wykładniczych grafów przypadkowych: a) klasycznego modelu blokowego, b) modelu blokowego z potęgowym rozkładem stopni węzłów. W obydwu sieciach, węzły należące do tego samego modułu oznaczono takim samym kolorem. Rozmiary węzłów są proporcjonalne do ich stopni, zaś grubość połączeń odpowiada roli, jaką pełnią one w procesach komunikacji międzywęzłowej całej sieci.



Rysunek 1.2: Przykładowe realizacje modułowych grafów wykładniczych o liczbie węzłów $N = 960$ oraz liczbie bloków $K = 6$; a) klasyczny model blokowy, b) model blokowy z potęgowym rozkładem oczekiwanych stopni węzłów.

1.4. Symulacje numeryczne metodą Monte Carlo

1.4.1. Wprowadzenie

W fizyce statystycznej, symulacje Monte Carlo (MC) umożliwiają numeryczne badanie układów fizycznych w stanie równowagi termodynamicznej¹⁷. Są one powszechnie wykorzystywane do modelowania układów i procesów zbyt złożonych, aby można było przewidzieć ich własności za pomocą podejścia analitycznego¹⁸.

Podstawową rolę w symulacjach MC odgrywa generowanie (losowanie) ciągu różnych realizacji badanego układu, np. ciągu grafów prostych o N wierzchołkach,

$$\{G_i\} = G_1, G_2, G_3, \dots \quad (1.42)$$

¹⁷Uwadze czytelnika polecamy doskonałą książkę M.E.J. Newman i G.T. Barkema [41] poświęconą metodom MC w fizyce statystycznej. O symulacjach MC w zastosowaniu do wykładniczych grafów przypadkowych można przeczytać w Dodatku E w książce A. i P. Fronczaków [3], oraz w pracach [11, 21]. W pierwszej z wymienionych prac, [11], szczegółowo omówiono algorytm Metropolis w zastosowaniu do klasycznych grafów przypadkowych. W drugiej pracy, [21], algorytm Metropolis został zaadoptowany do sieci o strukturze blokowej.

¹⁸Pouczającym przykładem może być tutaj znany w literaturze nt. sieci społecznych tzw. model Straussa [14, 15], o hamiltonianie: $H(G) = \alpha E(G) + \beta T(G)$ zależnym od liczby krawędzi, $E(G)$, i liczby trójkatów (tj. pętli o długości 3), $T(G)$, który został rozwiązany analitycznie dopiero w 2004 roku (zob. [42]), ponad 20 lat po pierwszych, głównie numerycznych pracach, w których dyskutowano jego własności.

Aby takie losowanie miało sens, możliwe realizacje badanego układu muszą się w tym ciągu pojawiać z zadaną częstością $P(G)$. Oczywiście, w badaniach wykładniczych grafów przypadkowych wspomniana częstość odpowiada prawdopodobieństwu realizacji grafu w zespole i jest opisana funkcją wykładniczą, $P(G) \propto e^{H(G)}$ (1.1).

Symulacje numeryczne polegające na bezpośrednim losowaniu ciągu (1.42) ze zbioru możliwych realizacji układu, $\mathcal{G} = \{G\}$, a następnie akceptowaniu tych realizacji z prawdopodobieństwem, $P(G)$, i odrzucaniu ich z prawdopodobieństwem, $1 - P(G)$, są (zazwyczaj) bardzo nieefektywne. Dzieje się tak z powodu (zazwyczaj) ogromnej liczby możliwych stanów badanego układu, z których większość ma bardzo małe prawdopodobieństwo realizacji $P(G) \ll 1$. Dlatego, w symulacjach Monte Carlo sekwencję (1.42) generuje się wykorzystując w tym celu ergodyczny proces Markowa¹⁹, w którym prawdopodobieństwo $p(G_i \rightarrow G_j)$ wygenerowania stanu (grafu) G_j ze stanu (grafu) G_i jest dobrane w taki sposób, by każdy z możliwych stanów badanego układu pojawiał się w tym ciągu z wymaganą częstością.

W dalszej części tego rozdziału omówimy warunki, jakie muszą spełniać prawdopodobieństwa przejść, $p(G_i \rightarrow G_j)$, aby można je było wykorzystać do utworzenia ciągu (1.42). Omówimy również znany algorytm Metropolis, który realizuje jedną z najprostszych postaci prawdopodobieństwa przejść.

1.4.2. Warunek równowagi szczegółowej

Jeśli badany układ jest w stanie równowagi, oznacza to, że charakteryzujący go rozkład prawdopodobieństwa $P(G)$ nie zależy od czasu. W odniesieniu do pojedynczego stanu G_i tego układu, warunek równowagi można zapisać w postaci równania:

$$\sum_{G_j \in \mathcal{G}} P(G_i) p(G_i \rightarrow G_j) = \sum_{G_j \in \mathcal{G}} P(G_j) p(G_j \rightarrow G_i), \quad (1.43)$$

które mówi o tym, że w stanie równowagi częstość przejść ze stanu G_i do dowolnego innego stanu G_j musi być równa częstości przejść w odwrotnym kierunku, tj. z dowolnego stanu G_j do G_i . Opuszczając sumy po obu stronach tego równania otrzymujemy tzw. warunek równowagi szczegółowej,

$$P(G_i) p(G_i \rightarrow G_j) = P(G_j) p(G_j \rightarrow G_i), \quad (1.44)$$

który w odniesieniu do wykładniczych grafów przypadkowych i po uwzględnieniu wykładniczej postaci rozkładu $P(G) \sim e^{H(G)}$ można przepisać jako:

$$\frac{p(G_i \rightarrow G_j)}{p(G_j \rightarrow G_i)} = \frac{P(G_j)}{P(G_i)} = e^{\Delta H}, \quad (1.45)$$

¹⁹Proces Markowa nazywamy ergodycznym, jeśli startując z dowolnego stanu badanego układu, w skończonej liczbie kroków, można dotrzeć do każdego innego stanu tego układu. W praktyce oznacza to, że podczas symulacji MC każda realizacja G układu, dla której $P(G) \neq 0$, musi mieć szansę na pojawienie się w ciągu (1.42).

gdzie

$$\Delta H = H(G_j) - H(G_i) \quad (1.46)$$

jest zmianą hamiltonianu badanego zespołu grafów, jaka nastąpiła w wyniku przejścia układu ze stanu G_i do stanu G_j .

1.4.3. Algorytm Metropolis

Prawdopodobieństwo przejścia $p(G_i \rightarrow G_j)$ występujące w warunku równowagi szczegółowej (1.45) można zapisać w postaci iloczynowej:

$$p(G_i \rightarrow G_j) = s(G_i \rightarrow G_j)a(G_i \rightarrow G_j), \quad (1.47)$$

gdzie $s(G_i \rightarrow G_j)$ reprezentuje prawdopodobieństwo selekcji stanu G_j , jeśli w ciągu (1.42) stanem poprzedzającym był stan G_i , zaś $a(G_i \rightarrow G_j)$ odpowiada prawdopodobieństwu akceptacji stanu G_j , tzn. umieszczeniu go we wspomnianym ciągu tuż za stanem G_i , tj. $\{\dots G_i, G_j \dots\}$. Jeśli przyjmiemy, że $s(G_i \rightarrow G_j)$ jest opisane rozkładem jednostajnym, wówczas warunek (1.45) uprości się do ilorazu prawdopodobieństw akceptacji:

$$\frac{a(G_i \rightarrow G_j)}{a(G_j \rightarrow G_i)} = e^{\Delta H}, \quad (1.48)$$

Algorytm Metropolis realizuje jedną z najprostszych postaci prawdopodobieństwa akceptacji dla zmiany stanów. Według tego algorytmu, jeśli prawdopodobieństwo $P(G_j)$ znalezienia układu w stanie G_j jest większe od $P(G_i)$, wówczas prawdopodobieństwo akceptacji przejścia $G_i \rightarrow G_j$ jest równe jedności:

$$a(G_i \rightarrow G_j) = 1 \quad \text{dla} \quad \Delta H > 0. \quad (1.49)$$

W przeciwnym wypadku jest ono opisane wykładniczą zależnością (1.48):

$$a(G_i \rightarrow G_j) = e^{\Delta H} \quad \text{dla} \quad \Delta H < 0. \quad (1.50)$$

1.4.4. Metoda Monte Carlo w wykładniczych grafach przypadkowych

Zastosowanie algorytmu Metropolis do badania wykładniczych grafów przypadkowych omówimy na przykładzie zespołu klasycznych grafów przypadkowych (zob. podrozdz. 1.3.1). Przypomnijmy, że hamiltonian tego zespołu, $H(G) = \theta E(G)$ (1.24), zależy jedynie od liczby krawędzi w grafie, $E(G)$, gdzie $\theta = \ln[p/(1-p)]$ (1.22) jest zewnętrznym parametrem zespołu, który można wyrazić poprzez prawdopodobieństwo p istnienia połączenia między dowolną parą węzłów.

W dalszej części tego podrozdziału założymy, że $\theta > 0$, tzn. omówimy wykorzystanie algorytmu Metropolis do grafów ER o prawdopodobieństwie połączenia

$p > 0,5$. Aby uprościć algorytm, dodatkowo przyjmujemy, że kolejne grafy w sekwencji (1.42), tj. $\{\dots G_i, G_j \dots\}$ będą się od siebie różniły jedynie pojedynczymi krawędziami, tzn. zmiana $G_i \rightarrow G_j$ będzie polegała na dodaniu lub usunięciu krawędzi w grafie G_i .

Przy tych założeniach możliwe zmiany hamiltonianu zespołu wynoszą:

$$\Delta H = \theta(E(G_j) - E(G_i)) = \pm\theta, \quad (1.51)$$

przy czym znak '+' odpowiada dodaniu, a '-' usunięciu krawędzi. Ponieważ jednak, zgodnie z założeniem, $\theta > 0$, to z równań (1.49) oraz (1.50) wynika, że zmianę grafu G_i polegającą na dodaniu krawędzi akceptujemy zawsze, zaś zmianę polegającą na usunięciu krawędzi akceptujemy z prawdopodobieństwem $e^{-\theta}$.

Podsumowując: Symulacja MC klasycznych grafów przypadkowych wykorzystująca algorytm Metropolis mogłaby przebiegać w następujący sposób:

- i. Symulację rozpoczynamy od dowolnego grafu prostego, np. od grafu pustego (tj. od pustej macierzy sąsiedztwa).
- ii. Następnie, losujemy dowolny element macierzy sąsiedztwa A_{ij} . Jeśli $A_{ij} = 0$ (co oznacza brak krawędzi między węzłami i oraz j), wtedy zastępujemy go przez $A_{ij} = 1$ (dodajemy krawędź). Jeśli natomiast $A_{ij} = 1$, wówczas z prawdopodobieństwem $e^{-\theta}$ zamieniamy²⁰ element macierzy sąsiedztwa: $A_{ij} = 0$ (co również oznacza, że z prawdopodobieństwem $(1 - e^{-\theta})$ element macierzy sąsiedztwa pozostaje niezmienny: $A_{ij} = 1$).
- iii. Po każdej zmianie stanu $G_i \rightarrow G_j$ (lub po każdych kolejnych n zmianach) sprawdzamy, jak w czasie zmienia się liczba krawędzi, $E(G)$. Gdy $\langle E \rangle = \text{const}$ oznacza to, że kolejne grafy w ciągu $\{\dots G_t, G_{t+1}, G_{t+2}, \dots\}$ pojawiają się z wykładniczą częstością, $P(G) \propto e^{H(G)}$.

W podobny sposób algorytm Metropolis można wykorzystać do badania dowolnych zespołów wykładniczych grafów przypadkowych, o dowolnych hamiltonianach (1.9).

1.5. Podsumowanie

Omówione w tej pracy, podejście do modelowania sieci złożonych wykorzystujące wykładnicze grafy przypadkowe umożliwia tworzenie sieci o dowolnych (z góry ustalonych) własnościach strukturalnych. W grafach tych, wspomniane własności strukturalne są w prosty sposób określone przez podanie tzw. hamiltonianu sieci. W tej

²⁰można to zaimplementować w następujący sposób: losujemy liczbę r z przedziału $[0, 1)$ (z rozkładu jednostajnego), a następnie sprawdzamy, czy $r < e^{-\theta}$;

pracy przedstawiono teoretyczne podstawy modelu, oraz trzy podstawowe przykłady zastosowania wykładniczych grafów przypadkowych: i. do badania klasycznych grafów przypadkowych, ii. w analizie sieci o zadanym rozkładzie stopni węzłów, oraz iii. do tworzenia modeli sieci o strukturze blokowej. Niestety, zabrakło miejsca i czasu, by omówić inne ważne modele wykładniczych grafów przypadkowych (np. znany w badaniach sieci społecznych model Straussa [14, 42]) i zastosowania tych grafów do analizy sieci rzeczywistych (np. do analizy sieci handlu światowego [16, 17]). Z myślą o tych czytelnikach, którzy podejście wykładniczych grafów przypadkowych chcieliby wykorzystać w swoich badaniach naukowych, w pracy umieszczono obszerny rozdział poświęcony symulacjom numerycznym Monte Carlo, w którym omówiono algorytm Metropolis i jego wykorzystanie do badania wykładniczych grafów przypadkowych.

Podziękowania

A.F. dziękuje Fundacji na rzecz Nauki Polskiej oraz Programowi Operacyjnemu Unii Europejskiej "Innowacyjna Gospodarka" za wsparcie finansowe na realizację projektu *Data mining, statistical analysis and predictive modelling in communication and information-driven complex networks* (nr projektu POMOST/2012-5/5).

Autor

Agata Fronczak

Wydział Fizyki, Politechnika Warszawska,
ul. Koszykowa 75, 00-662 Warszawa,
agatka@if.pw.edu.pl

Bibliografia

- [1] M.E.J. Newman, *Networks: an introduction*, Oxford University Press, Oxford, 2010.
- [2] S.N. Dorogovtsev, *Lectures on complex networks*, Oxford University Press, Oxford, 2010.
- [3] A. Fronczak, P. Fronczak, *Świat sieci złożonych: od fizyki do internetu*, Wydawnictwa Naukowe PWN, Warszawa, 2009.
- [4] B. Bollobás, *Random graphs*, Academic Press Inc., London Ltd., 1985.
- [5] A.-L. Barabási, R. Albert, *Emergence of scaling in random networks*, Science, 286:509, 1999.

- [6] A. Barrat, M. Barthélemy, R. Pastor-Satorras, A. Vespignani, *The architecture of complex weighted networks*, Proc. Natl. Acad. Sci. U.S.A. 101:3747–, 2004.
- [7] E. Ravasz, A.-L. Barabási, *Hierarchical organization in complex networks*, Phys. Rev. E 67:026112, 2003.
- [8] A. Lancichinetti, S. Fortunato, F. Radicchi, *Benchmark graphs for testing community detection algorithms*, Phys. Rev. E 78:046110, 2008.
- [9] C. Song, S. Havlin, H.A. Makse, *Self-similarity of complex networks*, Nature, 433:392, 2005.
- [10] C.J. Anderson, S. Wasserman, B. Crouch, *A p^* primer: logit models for social networks*, Social Networks, 21:37, 1999.
- [11] A. Fronczak, *Exponential random graph models*, rozdział w *Encyclopedia of social network analysis and mining*, R. Alhajj, J. Rpkne (Eds.), Springer-Verlag, 2014.
- [12] P.W. Holland, S. Leinhardt, *An exponential family of probability distributions for directed graphs*, J. Amer. Stat. Assoc. 76:33, 1981.
- [13] J.E. Besag, *Spatial interaction and the statistical analysis of lattice systems*, J. R. Stat. Soc. Ser. B (Stat. Methodol.) 36:192, 1974.
- [14] D. Strauss, *On a general class of models for interaction*, SIAM Rev. 28:513, 1986.
- [15] O. Frank, D. Strauss, *Markov graphs*, J. Am. Stat. Assoc. 81:832, 1986.
- [16] A. Fronczak, P. Fronczak, *Statistical mechanics of the international trade network*, Phys. Rev. E 85:056113, 2012.
- [17] A. Fronczak, *Structural hamiltonian of the international trade network*, Acta Phys. Pol. B. Proc. Supl. 5:31, 2012.
- [18] P. Erdős, A. Rényi, *On random graphs*, Publicationes Mathematicae, 6:290, 1959.
- [19] P. Erdős, A. Rényi, *On the evolution of random graphs*, Publications of the Mathematical Institute of the Hungarian Academy of Sciences, 5:17, 1960.
- [20] J. Park, M.E.J. Newman, *Statistical mechanics of networks*, Phys. Rev. E 70:066117, 2004.
- [21] P. Fronczak, A. Fronczak, M. Bujok, *Exponential random graph models for networks with community structure*, Phys. Rev. E 88:032810, 2013.
- [22] E.T. Jaynes, *Probability theory: the logic of science* G.L. Bretthorst (Ed.), Cambridge University Press, Cambridge, 2007.

- [23] J.N. Kapur, *Maximum-entropy models in science and engineering*, John Wiley & Sons, New York, 1989.
- [24] I. Ford, *Statistical thermodynamics: an entropic approach*, John Wiley & Sons Inc., 2013.
- [25] P. Attard, *Thermodynamics and statistical mechanics: equilibrium by entropy maximisation*, Academic Press, London, 2002.
- [26] E.T. Jaynes, *Information theory and statistical mechanics*, Phys. Rev. 106:620, 1957.
- [27] M.E.J. Newman, H.S. Strogatz, D.J. Watts, *Random graphs with arbitrary degree distributions and their applications*, Phys. Rev. E 64:026118, 2001.
- [28] R. Pastor-Satorras, A. Vespignani, *Epidemic spreading in scale-free networks*, Phys. Rev. Lett. 86:3200, 2001.
- [29] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Resilience of the Internet to random breakdowns*, Phys. Rev. Lett. 85:4626, 2000.
- [30] R. Cohen, K. Erez, D. ben-Avraham, S. Havlin, *Breakdown of the Internet under intentional attack*, Phys. Rev. Lett. 86:3682 (2001).
- [31] S. Fortunato, *Community detection in graphs*, Phys. Rep. 486:75, 2010.
- [32] D.J. Watts, S.H. Strogatz, *Collective dynamics of small world networks*, Nature 393:440, 1998.
- [33] A. Fronczak, P. Fronczak, J.A. Holyst, *Average path length in random networks*, Phys. Rev. E 70:056110, 2004.
- [34] R. Cohen, S. Havlin, *Scale-free networks are ultrasmall*, Phys. Rev. Lett. 90:058701, 2003.
- [35] M.E.J. Newman, *Communities, modules and large-scale structure in networks*, Nature Phys. 8:25, 2012.
- [36] M. Girvan, M.E.J. Newman, *Community structure in social and biological networks*, Proc. Natl. Acad. Sci. U.S.A. 99:7821, 2002.
- [37] M.E.J. Newman, M. Girvan, *Finding and evaluating community structure in networks*, Phys. Rev. E 69:026113, 2004.
- [38] B. Karrer, M.E.J. Newman, *Stochastic blockmodels and community structure in networks*, Phys. Rev. E 83:016107, 2011.
- [39] P.W. Holland, K.B. Laskey, S. Leinhardt, *Stochastic blockmodels: first steps*, Social Networks, 5:109, 1983.
- [40] C. Song, S. Havlin, H.A. Makse, *Origins of fractality in the growth of complex networks*, Nature Phys. 2:275, 2006.

- [41] M.E.J. Newman, G.T. Barkema, *Monte Carlo methods in statistical physics*, Clarendon Press, Oxford, 1999.
- [42] J. Park, M.E.J. Newman, *Solution of the two-star model of a network*, Phys. Rev. E 70:066146, 2004.