

PAWEŁCZYK Marek¹
SKROBACKI Zbigniew²

Sieci złożone jako narzędzie modelowania systemów transportowych i logistycznych

WSTĘP

Praca poświęcona jest charakterystyce problematyki modelowania sieci złożonych, mogących znaleźć zastosowanie przy opisie i badaniach różnego rodzaju systemów funkcjonujących we współczesnym świecie, w tym systemów transportowych i logistycznych. Przedstawiono wyniki badań ukierunkowanych na modele sieci ewoluujących, które stanowią aktualnie najlepsze przybliżenie opisu sieci rzeczywistych.

W literaturze angielskojęzycznej, ukierunkowanej na badania sieci, upowszechnił się już termin *network science*, tłumaczony na język polski jako nauka o sieciach [1]. Z uwagi na ścisłe powiązania struktury sieci z jej cechami funkcjonalnymi modelowanie sieci nie służy jedynie celom poznawczym - dzięki uproszczonym modelom można w coraz większym stopniu przewidywać zachowanie i rozwój sieci w różnych sytuacjach. Znajomość podstawowych reguł ewolucji sieci daje także możliwość wpływu na nią - np. znajomość struktury sieci zależności pokarmowych w pewnym ekosystemie ułatwia ochronę zagrożonych gatunków, a znajomość reguł, określających rozwój Internetu, stanowi warunek efektywnego projektowania wydajnych protokołów sieciowych. Już obecnie opracowane modele teoretyczne sieci WWW są wykorzystywane do tworzenia oraz testowania wyszukiwarek internetowych [1].

Powszechny jest pogląd, że pierwszą pracą, związaną z problematyką sieci, jest artykuł L. Eulera, poświęcony tzw. zagadnieniu mostów królewieckich (1736). Rozwiązanie tego problemu jest uważane za pierwsze twierdzenie teorii grafów, szeroko wykorzystywanej w opisie sieci. Teoria grafów rozwinęła się w II poł. XIX w. dzięki pracom J. Sylwestera, będącego twórcą pojęcia graf - ang. *graph*) [3]. W XX wieku grafy zaczęto stosować w socjologii - w 1933 roku psycholog i socjolog J. Moreno do opisu struktury społecznej grupy uczniów zastosował tzw. socjogram - graficzną prezentację relacji, występujących pomiędzy członkami grupy. Socjogramy znalazły szereg zastosowań w analizie tzw. sieci społecznych (ang. *social network*) [4].

Prace P. Erdősa i A. Rényiego, poświęcone tzw. grafom losowym [5], stanowiły przełom w rozwoju nauki o sieciach. W roku 1998 D. Krakchardt i K. Carley wprowadzili pojęcie meta-sieci z modelem PCANS (*Precedence, Commitment of resources, Assignment of individuals to tasks, Networks of relations among personnel and Skills linking individuals to resources* - pierwszeństwo, zaangażowanie zasobów, przydział jednostek do zadań, sieci relacji między pracownikami i umiejętności wiążące jednostki z zasobami). Pojęcie to stanowi obecnie podstawę koncepcji sieci obejmujących różne domeny powiązane ze sobą i dało podwaliny do powstania dyscypliny nauki o sieciach, określanej jako analiza dynamiki sieci (ang. *dynamic network analysis*). Badania, związane z nauką o sieciach, koncentrują się w ostatnim okresie wokół opisu matematycznego różnych topologii sieciowych. M.in. model zaproponowany przez D. Watta i S. H. Strogatza (tzw. *small-world network* - sieci małych światów) umożliwił uzyskanie dużej zgodności pomiędzy danymi empirycznymi a ich reprezentacją matematyczną [6]. A.-L. Barabási i R. Albert odkryli zjawisko bezskalowości sieci (ang. *scale-free network*) - właściwość sieci, oznaczającą jej topologię, charakteryzującą się dużym, spójnym komponentem (*hub*) z licznymi połączeniami, rosnącym

¹ Politechnika Świętokrzyska w Kielcach, Wydział Zarządzania i Modelowania Komputerowego, Al. Tysiąclecia Państwa Polskiego 7, 25-414 Kielce; Tel. +48 41 342-46-96, Fax: +48 41 342-43-06, m.pawelczyk@tu.kielce.pl

² Politechnika Świętokrzyska w Kielcach, Wydział Mechatroniki i Budowy Maszyn; Al. Tysiąclecia Państwa Polskiego 7, 25-414 Kielce; Tel. +48 41 342-44-48, Fax: +48 41 344-29-97, zbig@tu.kielce.pl

(w sieciach ewoluujących) w taki sposób, aby została utrzymana stałość relacji liczby połączeń do liczby węzłów.

Praca stanowi próbę nakreślenia ogólnej charakterystyki sieci – jako obiektu badań naukowych oraz wskazania na możliwe obszary zastosowań tej teorii, szczególnie w obrębie problematyki transportowo-logistycznej.

1. ELEMENTY TEORII GRAFÓW

Graf $G(N,E)$ definiuje się jako zbiór N wierzchołków (węzłów), połączonych E krawędziami. Zbiór wierzchołków (niepusty i skończony) grafu G oznacza się jako $V(G)$, zaś zbiór krawędzi, stanowiących różne, nieuporządkowane pary wierzchołków, należących do V oznacza się $E(G)$. Za pomocą wierzchołka można modelować np. węzeł w sieci drogowej, a za pomocą krawędzi – relację między nimi (np. drogę łączącą dwa węzły drogowe). Powyższa definicja określa grafy proste (nieskierowane), stanowiące najprostszą klasę grafów. Gdy wszystkim krawędziom grafu zostaną przyporządkowane kierunki (dowolna para wierzchołków połączonych krawędzią będzie uporządkowana), to graf określa się jako skierowany (digraf) – w takiej sytuacji zachodzi jednokierunkowa relacja między wierzchołkami (np. w sieci cytowań pewien artykuł, reprezentowany przez wierzchołek a grafu został zacytowany w artykule, reprezentowanym przez wierzchołek b). Multigrafem jest graf, w którym mogą występować krawędzie wielokrotne (dwa wierzchołki mogą być połączone więcej niż jedną krawędzią). Można też przypisać każdej krawędzi pewną liczbę (wagę), określającą pewną własność krawędzi (np. maksymalną przepustowość odcinka drogi). Grafy z taką własnością określa się jako grafy ważone. Specjalną klasą grafów są grafy z wierzchołkami należącymi do dwóch rozłącznych zbiorów i krawędziami łączącymi jedynie węzły należące do jednego zbioru z węzłami należącymi do drugiego zbioru węzłów – są to tzw. grafy dwudzielne (na przykład sieć obejmująca zbiór autorów książek i zbiór tytułów książek).

Strukturę grafu przedstawia się za pomocą macierzy kwadratowej o wymiarach $N \times N$, zwanej macierzą sąsiedztwa A (ang. adjacency matrix). Elementy a_{ij} macierzy definiują sposób połączenia wierzchołka i z wierzchołkiem j : w grafie prostym macierz A jest symetryczna i zawiera elementy równe 1 (krawędź i - j istnieje) lub 0 (brak krawędzi). W grafach skierowanych Macierz A jest asymetryczna. W przypadku multigrafów wartości a_{ij} oznaczają liczbę krawędzi łączących wierzchołek i z wierzchołkiem j . Elementy na głównej przekątnej macierzy A są równe zeru, za wyjątkiem sytuacji, gdy graf zawiera pętle (krawędzi łączące i -ty wierzchołek grafu z nim samym). W grafach ważonych elementy macierzy sąsiedztwa są wagą konkretnej krawędzi.

Stopień węzła i sieci (wierzchołka i grafu) – k_i – to liczba bezpośrednich połączeń węzła i z innymi wierzchołkami sieci. Wierzchołki sieci, połączone z danym wierzchołkiem krawędzią, określane są jako pierwsi (najbliżsi) sąsiedzi, a zatem stopień węzła sieci jest równy liczbie najbliższych sąsiadów. W grafie prostym stopień węzła i jest równy sumie elementów macierzy sąsiedztwa (liczbie niezerowych elementów macierzy) w kolumnie/wierszu i :

$$k_i = \sum_{j=1}^N a_{ij} = \sum_{j=1}^N a_{ji} \quad (1)$$

Stąd liczba wszystkich krawędzi w sieci:

$$L = \frac{1}{2} \sum_{i=1}^N k_i \quad (2)$$

i przeciętny (średni) stopień wierzchołka:

$$\langle k \rangle = \frac{2L}{N}. \quad (3)$$

W przypadku grafów (i sieci) skierowanych stopnie wejściowe węzła k^{in} są w ogólności różne od stopni wyjściowych k^{out} , przez analogię definiuje się liczbę krawędzi wejściowych K^{in} i wyjściowych K^{out} oraz średni stopień wierzchołka wejściowego $\langle k_{in} \rangle$ i wyjściowego $\langle k_{out} \rangle$. Stopień wejściowy wierzchołka i uzyskuje się sumując elementy macierzy sąsiedztwa wzdłuż kolumn, a stopień wyjściowy – sumując a_{ij} wzdłuż i -tego wiersza macierzy.

$$k_i^{in} = \sum_{j=1}^N a_{ij}; \quad k_i^{out} = \sum_{j=1}^N a_{ji} \quad (4)$$

W sieciach ważonych, w których poszczególnym krawędziom przypisane są wagi, definiuje się stopień ważony, zwany także **siłą** wężła:

$$s_i = \sum_{j=1}^N a_{ij} \quad (5)$$

Ciąg wierzchołków i krawędzi, zestawionych w taki sposób, że każdy kolejny wierzchołek sąsiaduje z wierzchołkiem go poprzedzającym i z wierzchołkiem bezpośrednio po nim następującym określa się jako trasę w grafie G :

$$\langle i, \{i, v_1\}, v_1, \{v_1, v_2\}, v_2, \dots, v_n, \{v_n, j\}, j \rangle. \quad (6)$$

Trasa, w której wszystkie krawędzie są różne – to **ścieżka**. W przypadku, gdy początek ścieżki pokrywa się z jej końcem, tzn. $i = j$, ścieżkę nazywa się **cyklem**. **Drogą** w grafie G jest ciąg krawędzi $\{\{v_1, v_2\}, \{v_2, v_3\}, \dots, \{v_{n-1}, v_n\}\}$, spełniających warunek, że koniec krawędzi $\{v_i, v_{i+1}\}$ jest początkiem krawędzi $\{v_{i+1}, v_{i+2}\}$ dla każdego $i = 0, \dots, n-2$, przy czym wszystkie krawędzie i wszystkie wierzchołki są różne. Początek krawędzi $\{v_1, v_2\}$ - to początek drogi, a koniec krawędzi $\{v_{n-1}, v_n\}$ – to koniec drogi. **Długością** drogi jest liczba tworzących ją krawędzi. Dowodzi się, że każdy element (i, j) x -tej potęgi macierzy sąsiedztwa – A^x jest równy liczbie różnych tras o długości x pomiędzy wierzchołkami i oraz j . **Odległością** $d(i, j)$ wierzchołka i od wierzchołka j jest długość najkrótszej drogi prowadzącej z i do j , lub najmniejsza liczba krawędzi, łączących wierzchołek i z wierzchołkiem j . Wierzchołki odległe od siebie o x krawędzi nazywa się x -tymi sąsiadami.

Graf nazywa się **spójnym**, gdy dla każdej pary wierzchołków i oraz j można wyznaczyć drogę łączącą te dwa wierzchołki. Odległość pomiędzy jego dwoma dowolnymi wierzchołkami i oraz j grafu spójnego jest najmniejszą liczbą x , dla której element $a_{ij}^{(x)}$ macierzy A^x jest różna od zera.

Graf niespójny posiada wierzchołki izolowane i składa się z pewnej liczby **składowych spójnych (komponentów, klasterów)**. W dużych sieciach, z którymi najczęściej mamy do czynienia (np. Internet), większość węzłów należy do jednego, wielkiego, spójnego komponentu, który można uzyskać zaniedbując kierunki krawędzi i znajdując w powstałej w ten sposób sieci nieskierowanej największy spójny komponent. Wewnątrz niego można wyróżnić [1]:

- największy silnie spójny komponent (NSSK) - rdzeń sieci skierowanej, w którym, przemieszczając się wzdłuż krawędzi można dotrzeć z dowolnego węzła do każdego innego węzła w NSSK;
- komponent wejściowy, zawierający wszystkie węzły, od których można znaleźć drogę do NSSK; z powyższego wynika, że NSSK jest podzbiorem komponentu wejściowego;
- komponent wyjściowy, obejmujący wszystkie węzły, do których można dotrzeć z NSSK – a zatem NSSK jest podzbiorem komponentu wyjściowego;
- wąsy - te węzły, do których nie można dotrzeć z NSSK oraz te, z których nie można dotrzeć do NSSK;
- komponenty izolowane.

Pojęcie spójności sieci (i grafu) ma bezpośredni związek z fizyczną teorią perkolacji [10] i z teorią grafów losowych. Z zagadnieniami spójności w grafach prostych wiąże się ważne twierdzenie o liczbie krawędzi grafu [9], które głosi, że jeżeli graf prosty G z liczbą wierzchołków równą N ma t składowych (spójnych podgrafów grafu G nie zawartych w większym podgrafie spójnym grafu G), to liczba E jego krawędzi spełnia nierówność:

$$(N-t) \leq E \leq (N-t)(N-t+1)/2. \quad (7)$$

Duże znaczenie w teorii grafów mają pojęcia: las i drzewo. Las - to graf prosty, który nie zawiera cykli, zaś drzewo – to las spójny. Drzewo o N wierzchołkach posiada następujące własności: nie zawiera cykli i ma $N-1$ krawędzi; każda jego krawędź jest mostem - krawędzią, której usunięcie powoduje, że graf przestaje być spójny; każde dwa wierzchołki grafu, będącego drzewem, są połączone dokładnie jedną drogą; przez dodanie dowolnej nowej krawędzi do drzewa otrzymuje się graf z dokładnie jednym cyklem [9].

W sieciach złożonych dużą rolę odgrywa parametr charakteryzujący własności sieci - współczynnik gronowania (klasteryzacji - ang. clustering coefficient). Termin ten w najprostszej wersji dotyczy grafów prostych i jest miarą charakteryzującą liczbę krawędzi w otoczeniu badanego wierzchołka. Jest to iloraz liczby L_i istniejących krawędzi między sąsiadami wierzchołka do liczby wszystkich możliwych krawędzi między tymi sąsiadami, która wynosi $\frac{1}{2} k_i(k_i - 1)$. Stąd:

$$C_i = \frac{2L_i}{k_i(k_i - 1)}. \quad (8)$$

przy czym $0 \leq L_i \leq k_i(k_i - 1)/2$, i w związku z tym $0 \leq C_i \leq 1$. Współczynnik gronowania opisuje zatem prawdopodobieństwo, że pierwsi (najbliżsi) sąsiedzi wężła i są względem siebie również pierwszymi sąsiadami [7].

Istotną rolę w analizie sieci, szczególnie sieci transportowych i logistycznych, odgrywa macierz sąsiedztwa, która dla grafów prostych (nieskierowanych) jest symetryczna, co wynika z jej definicji. Dowodzi się, że ma ona wszystkie wartości własne rzeczywiste i pełen zbiór ortogonalnych wektorów własnych. Zbiór wartości własnych tej macierzy określa się jako widmo grafu.

Przez permutację macierzy sąsiedztwa otrzymuje się grafy izomorficzne (o strukturze, która może być uważana za identyczną). Grafy izomorficzne posiadają ten sam wielomian charakterystyczny, zbiór wartości własnych, wyznacznik oraz ślad (tzn. sumę elementów na głównej przekątnej macierzy); nie jest prawdziwa zależność odwrotna - grafy z takim samym wielomianem charakterystycznym nie muszą być izomorficzne.

W przypadku grafów skierowanych macierz $I - A$ (gdzie I - jest macierzą jednostkową) posiada macierz odwrotną wtedy i tylko wtedy, gdy graf G nie zawiera cykli. Elementy (i, j) macierzy $(I - A)^{-1}$ są równe liczbie wszystkich ścieżek z wierzchołka i do wierzchołka j (przy braku cykli ta liczba jest skończona), co bezpośrednio wynika z rozwinięcia tej odwrotności w szereg geometryczny:

$$(I - A)^{-1} = I + A + A^2 + A^3 + \dots \quad (9)$$

i sumowania liczby ścieżek o długości 0, 1, 2 itd.

Macierz sąsiedztwa grafu dwudzielnego, z dwoma grupami wierzchołków o liczebności r i s ma postać:

$$A = \begin{bmatrix} O & B \\ B^T & O \end{bmatrix}, \quad (10)$$

gdzie: B jest macierzą o wymiarach $r \times s$, a O jest macierzą zerową. Oznacza to, że macierz B w pełni opisuje własności grafu dwudzielnego.

Suma stopni wierzchołków grafu, K , jest równa podwojonej łącznej liczbie krawędzi, L :

$$K = 2L = \sum_{i,j} a_{ij} = \sum_i (A^2)_{ii} = Tr A^2 \quad (11)$$

przy czym symbol $Tr A^2$ oznacza ślad macierzy A^2 (sumę jej elementów diagonalnych). Łączna liczba cykli stopnia 3 w grafie wynosi [7]:

$$N_3 = \frac{1}{6} \sum_i (A^3)_{ii} = \frac{1}{6} Tr A^3, \quad (12)$$

a liczba połączonych trójek wierzchołków w grafie jest równa [7]:

$$T = \frac{1}{2} \sum_{i \neq j} (A^2)_{ij} = \frac{1}{2} \sum_i k_i(k_i - 1). \quad (13)$$

W literaturze [7] można ponadto znaleźć zależność pozwalającą na wyznaczenie współczynnika gronowania całego grafu jako funkcji macierzy sąsiedztwa:

$$C = \frac{1}{9} \frac{\sum_i (A^3)_{ii}}{\sum_{i \neq j} (A^2)_{ij}} \quad (14)$$

Ponadto na podstawie macierzy sąsiedztwa definiuje się długość najkrótszej drogi ℓ pomiędzy wierzchołkami i oraz j : jest to najniższy wykładnik potęgowy ℓ , dla którego element (i, j) jest różny od zera:

$$(A^{\ell-1})_{ij} = 0; \quad (A^{\ell})_{ij} \neq 0. \quad (15)$$

2. STATYCZNE CHARAKTERYSTYKI SIECI

Sieci różnią się między sobą **rozkładem stopni wierzchołków** definiowanym jako rozkład prawdopodobieństwa $P(k)$. Ogólnie wielkość $P(k)$ określa udział węzłów w sieci o stopniu k , innymi słowy prawdopodobieństwo, że losowo wybrany węzeł będzie miał stopień k .

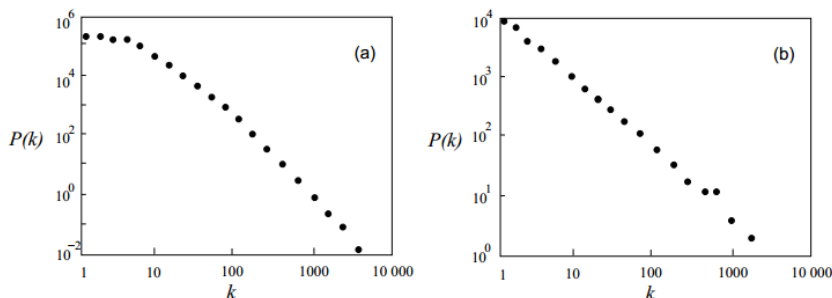
Konsekwencją uproszczenia układów rzeczywistych do zbioru węzłów i krawędzi (a więc ich prezentacji w postaci sieci) jest utrata informacji, z drugiej strony można dzięki temu zastosować te same narzędzia do analizy całkiem odmiennych układów. Badania, prowadzone pod koniec ubiegłego wieku przez Alberta-László Barabásiego doprowadziły do odkrycia nieznanej dotąd, ale powszechnie występującej w sieciach rzeczywistych własności: okazało się, że rozkład stopni wierzchołków w sieciach rzeczywistych, ma charakter potęgowy, a zatem:

$$P(k) = C \cdot k^{-\alpha}, \quad \alpha > 0. \quad (16)$$

W konsekwencji rozkład potęgowy nie ma naturalnej skali, mówi się, że jest „bezskałowy”, co oznacza, że wielu przypadkach posługiwanie się wielkościami średnimi prowadzi do istotnych błędów. W sieciach takich wiele węzłów ma tylko jedną krawędź, jednocześnie występują węzły z ogromną liczbą krawędzi, tzw. huby. Dobrym przykładem takich sieci jest sieć połączeń lotniczych - w której występuje niewielka liczba wielkich węzłów lotniczych oraz bardzo wiele słabo usieciowionych lotnisk o małej liczbie połączeń [1]. Rozkłady takie określa się jako „rozkłady z tłustym ogonem” (ang. fat-tailed distribution). Wyniki badań, prowadzonych przez różnych autorów, opisane m.in. w [7] wskazują na fakt, że nie wszystkie sieci „tłustoogonowe” są sieciami potęgowymi – najłatwiej się o tym można przekonać sporządzając wykres zależności $P(k)$ w skali podwójnie logarytmicznej. Logarytmując zależność (16) otrzymamy:

$$\ln P(k) = \ln C - \alpha \ln k. \quad (17)$$

Wynika stąd, że zależność potęgowa (16) we współrzędnych $\ln k$, $\ln P(k)$ staje się równoważna funkcji liniowej.



Rys. 1. Rozkład empiryczny stopni wierzchołków dwóch sieci rzeczywistych współautorów publikacji a) baza Medline b) baza SPIRES [11]

Jak widać na rys. 1 rozkład stopni wierzchołków sieci współautorów w bazie Medline nie może być aproksymowany prostą zależnością potęgową (16), w przeciwieństwie do sieci współautorów w bazie SPIRES, charakteryzującej się rozkładem potęgowym o wykładniku $\alpha \approx 1,2$ [11].

W pracy [7] przedstawiono przegląd prac ukierunkowanych na wyznaczenie współczynnika gronowania dla sieci rzeczywistych. Wartości współczynnika C dla dużych sieci wynosiły od 0,066 dla sieci współautorstwa w bazie Medline do 0,79 dla sieci aktorów występujących wspólnie w filmach. Współczynnik gronowania dla sieci WWW oszacowano na 0,11, a dla Internetu na poziomie domen – w przedziale 0,18 – 0,3.

Pojęcie „sieć małego świata (ang. small world network) pojawiło się w związku z eksperymentem, przeprowadzonym w 1967 roku przez amerykańskiego socjologa S. Milgrama, w ramach którego wysłano 160 listów do losowo wybranych osób, z prośbą o przekazanie listu do osoby, której zdjęcie, nazwisko i adres dołączono do listu. Adresat mógł przesłać list bezpośrednio do osoby wymienionej w liście, o ile znał ją osobiście, albo przesłać go do znajomego, o którym mógł sądzić, że zna poszukiwaną osobę lub zna kogoś, kto ją zna. Celem eksperymentu było wyznaczenie długości łańcucha znajomych, gwarantującego dostarczenie przesyłki do adresata. 42 listy dotarły do poszukiwanej osoby, przy czym w kilku przypadkach wystarczyło zaledwie dwóch pośredników. Średnia wyniosła sześciu pośredników, w kilku skrajnych przypadkach pośredników było aż kilkunastu [9]. Milgram wykazał, że w sieci społecznej, obejmującej kilka miliardów ludzi, przeciętna droga, łącząca dwa węzły, składa się mniej więcej z sześciu kroków. Pojawiło się sformułowanie „sześć stopni separacji” lub „sześć uścisków dłoni” - od dowolnej osoby na świecie dzieli nas przeciętnie sześć uścisków dłoni.

W czerwcu 2006 roku Amerykanie E. Horvitz i J. Leskovec przeprowadzili analizę 30 mld wiadomości, wysłanych z 240 mln komputerów. Przeciętna droga pomiędzy dwoma dowolnymi użytkownikami składała się z około 6,6 ogniw pośrednich.

Jak widać, sieci społeczne mogą być nazwane małymi światami: przeciętna długość drogi pomiędzy dwoma dowolnymi węzłami sieci jest w mała w porównaniu z liczbą węzłów, mimo że liczba krawędzi nie jest znacznie większa od liczby wierzchołków.

Kolejnym, ważnym pojęciem w teorii sieci złożonych, jest pojęcie centralności (ang. centrality), opisującej względną ważność węzła w grafie. Jest to także ocena intuicyjnego odczucia że w sieci niektóre wierzchołki lub krawędzie są usytuowane bardziej centralnie (są ważniejsze) niż inne. Centralność definiuje się za pomocą kilku miar, spośród których najważniejsze cztery - to stopień centralności C_D (ang. degree centrality), centralność sąsiedztwa C_C (ang. closeness centrality), pośrednictwo C_B (ang. betweenness) oraz centralność wektora własnego (eigenvector centrality) [12]. Najstarsze i najprostsze pojęcie – to **centralność stopnia**, definiowana poprzez liczbę połączeń (krawędzi), którą sieć posiada. W przypadku sieci z połączeniami skierowanymi oddzielnie definiuje się miary dla krawędzi wejściowych i wyjściowych (indegree i outdegree). W sieciach społecznych indegree stanowi miarę popularności (popularity), a outdegree - towarzyskości (gregariousness). Centralność stopnia dla wierzchołka (węzła) v w grafie $G = (V, E)$ definiuje się jako:

$$C_D(v) = \deg(v) \quad (18)$$

Niech v^* będzie węzłem z najwyższą centralnością stopnia w grafie G , a $X = (Y, Z)$ będzie grafem z wierzchołkami połączonymi krawędziami ze zbioru Y , maksymalizującym wartość wyrażenia:

$$H = \sum_{j=1}^{|Y|} C_D(y^*) - C_D(y_j), \quad (19)$$

gdzie y^* oznacza węzeł z najwyższą centralnością stopnia w X .

Centralność stopnia dla całego grafu G będzie równa:

$$C_D(G) = \frac{\sum_{i=1}^{|V|} [C_D(v^*) - C_D(v_i)]}{H} \quad (20)$$

Wartość H jest największa, gdy graf X zawiera jeden centralny węzeł (wierzchołek), do którego dołączone są wszystkie pozostałe węzły. Wtedy $H = (n - 1)(n - 2)$.

Naturalną miarą, określającą odległość pomiędzy poszczególnymi parami węzłów w grafie jest długość najkrótszej drogi łączącej te węzły - oddaloność (ang. farness), którą dla węzła s definiuje się jako sumę odległości do wszystkich pozostałych węzłów grafu, a centralność sąsiedztwa – jako odwrotność oddaloności [13]. Im bardziej centralnie położony jest węzeł, tym mniejsza jest jego łączna odległość do innych węzłów. Danganchev określa centralność sąsiedztwa w taki sposób, aby można było w tej definicji uwzględnić również grafy, w których nie występują połączenia pomiędzy częścią węzłów. Zgodnie z [14]:

$$C_C(v) = \sum_{t \in V \setminus v} 2^{-d_G(v,t)} \quad (21)$$

Miarą centralności wężła w grafie, względnie miarą centralności krawędzi w grafie jest pośrednictwo (ang. betweenness). W przypadku miary centralności wężła miara ta opiera się na założeniu, że komunikacja odbywa się tylko wzdłuż najkrótszych dróg. Jeśli symbolem δ_{jk} oznaczy się liczbę najkrótszych dróg, łączących wężły j oraz k , a przez $\delta_{jk}^{(i)}$ - liczbę dróg, przechodzących przez wierzchołek i , to miarą pośrednictwa wężła i będzie uśredniony iloraz tych dwóch wielkości [15]:

$$C_{Bi}^V = \frac{2}{(N-1)(N-2)} \sum_k \sum_{j>k} \frac{\delta_{jk}^{(i)}}{\delta_{jk}}, \quad (22)$$

przy czym $j, k \neq i$.

Pośrednictwo charakteryzuje odporność sieci na uszkodzenia wężła - określa, ile najkrótszych dróg zniknie w grafie w przypadku uszkodzenia i -tego wężła sieci. W przypadku usunięcia wężłów, których miara pośrednictwa jest największa, pojawią się największe zakłócenia działania sieci. Można także wyznaczyć wrażliwość sieci na usunięcie krawędzi - w takiej sytuacji definiuje się pośrednictwo krawędzi e zgodnie z zależnością:

$$C_{Bi}^E = \frac{2}{N(N-1)} \sum_k \sum_{j>k} \frac{\delta_{jk}^{(e)}}{\delta_{jk}}, \quad (23)$$

przy czym $\delta_{jk}^{(e)}$ oznacza liczbę najkrótszych dróg przechodzących przez krawędź e . Za pomocą tej miary można np. analizować wrażliwość sieci drogowej na zablokowanie jednego z połączeń.

Centralność wektora wartości własnych (ang. eigenvector centrality) jest wskaźnikiem określającym istotności (wpływ) wężła w sieci. Wskaźnik ten przyporządkowuje względne miary wszystkim wężłom w sieci, opierając się na założeniu, że połączenia z wężłami ocenianymi wyżej mają większy udział we wskaźniku, niż połączenia z wężłami ocenianymi niżej. Wariantem tego wskaźnika jest stosowana przez firmę Google miara PageRank [16].

Do wyznaczania wskaźnika można wykorzystać macierz sąsiedztwa A . Niech $A = [a_{vt}]$ będzie macierzą sąsiedztwa grafu $G = (V, E)$, którego liczba wężłów wynosi $|V|$. Zakładamy, że gdy wężel v jest połączony z wężłem t , to $a_{vt} = 1$, w przeciwnym przypadku $a_{vt} = 0$. Centralność wektora własnego dla wężła v można zdefiniować jako:

$$x_v = \frac{1}{\lambda} \sum_{t \in M(v)} x_t = \frac{1}{\lambda} \sum_{t \in G} a_{vt} x_t \quad (24)$$

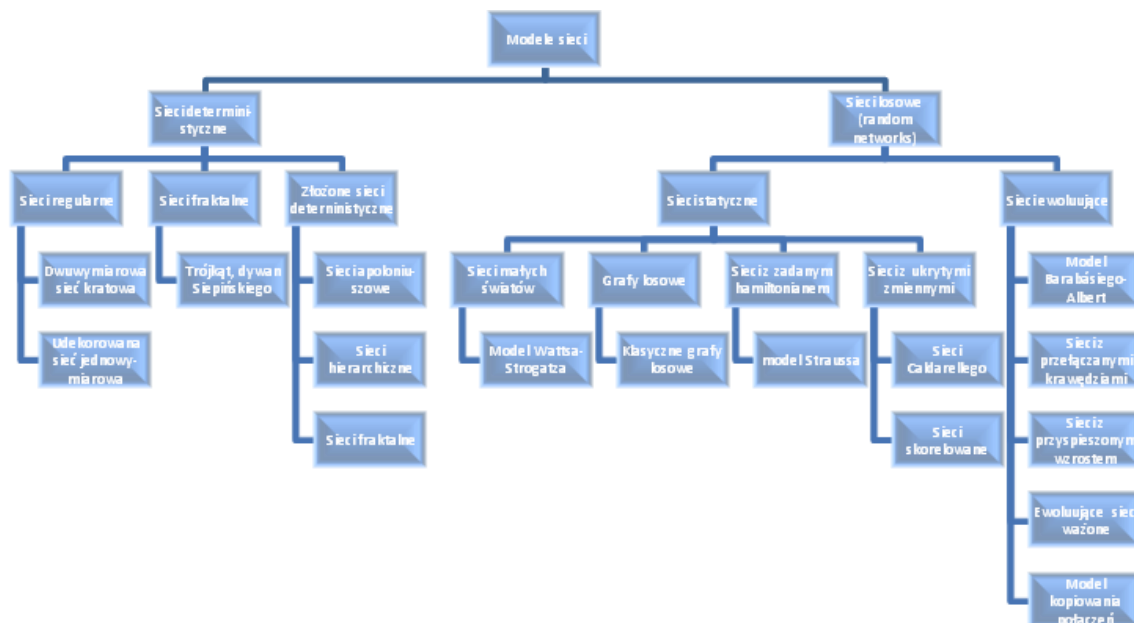
przy czym $M(v)$ - to zbiór sąsiadów wężła v , a λ jest stałą (skalarem), którą można wyznaczyć z równania macierzowego.

$$A \mathbf{x} = \lambda \mathbf{x} \quad (25)$$

Rozwiązanie równania (25) daje w ogólnym przypadku wiele różnych wartości własnych λ , dla których istnieje wektor własny. Z wymogu, aby wszystkie wartości wektora własnego były dodatnie, wynika konieczność wyboru największej wartości własnej λ we wskaźniku (24) [17]. v -ta składowa wektora wartości własnych stanowi miarę centralności wektora wartości własnych. Opisana metoda wyznaczania wskaźnika może być uogólniona dla macierzy sąsiedztwa dla grafów ważonych.

3. MODELE SIECI Z GRAFAMI LOSOWYMI

Najbardziej elementarną klasyfikacją modeli sieci złożonych jest podział na sieci **deterministyczne** i **losowe** (zwane także sieciami przypadkowymi). W sieciach losowych wierzchołki są połączone krawędziami w sposób losowy - co oznacza, że prawdopodobieństwo połączenia dwóch losowo wybranych wierzchołków wynosi p . Należy zaznaczyć, że mówiąc o modelu sieci losowej rozważa się nie jeden, konkretny obiekt, lecz całą przestrzeń jej możliwych realizacji. Jest to zasadnicza cecha odróżniająca sieci losowe od deterministycznych, w przypadku których nie dopuszcza się żadnej przypadkowości [1].



Rys. 2. Klasyfikacja modeli sieci. Opracowanie własne na podstawie [9]

Sieci losowe (zwane także sieciami przypadkowymi) dzieli się na sieci (konstrukcje) **statyczne** i sieci **ewoluujące**, przez analogię do podziału fizyki statystycznej na równowagową (odnoszącą się do układów w stanie równowagi) oraz nierównowagową (obejmującą układy niebędące w stanie równowagi, ale mogące przebywać w pewnym nierównowagowym stanie stacjonarnym) [9]. Makroskopowe charakterystyki sieci równowagowych (liczba węzłów i połączeń, współczynnik gronowania, rozkład stopni wierzchołków) nie zależą od czasu. W przypadku sieci ewoluujących ich charakterystyki (m.in. liczba węzłów i krawędzi, wzorce połączeń międzywęzłowych), ulegają zmianom w czasie, choć zmiany te zazwyczaj mają cechy nierównowagowych procesów stacjonarnych.

Podstawą przedstawionej na rys. 2 klasyfikacji modeli sieci jest zagadnienie mechanizmów wzrostu i zmienności badanych układów. Przykładem sieci ewoluujących może być, na przykład, sieć WWW.

Modele sieci statycznych można stosować do modelowania procesów dynamicznych, występujących w sieciach rzeczywistych, w których mechanizmy wzrostu sieci nie są istotne (np. modele epidemii). W takim przypadku rozważa się własności strukturalne sieci, np. rozkład potęgowej stopni wierzchołków i ich wpływ na funkcjonalność rozważanego układu.

Klasyczne grafy przypadkowe po raz pierwszy opisała dwójka matematyków – P. Erdős i A. Rényi w 1960 roku. Graf losowy $G_{N,E}$, a dokładniej zbiorowość statystyczną takich grafów $\{G_{N,E}\}$, definiuje procedura, składająca się z dwóch etapów:

- najpierw ustala się liczbę węzłów N , należących do grafu i wyznacza te węzły,
- następnie dokonuje się losowego wyboru $|E|$ par spośród wszystkich $\binom{N}{2}$ par węzłów i tworzy między nimi połączenia.

W uogólnionym modelu klasycznym parametr $|E|$, opisujący liczbę krawędzi, zastąpiono prawdopodobieństwem p występowania krawędzi między dowolną parą wierzchołków. W zmodyfikowanej procedurze budowy grafów w etapie drugim przypisuje się z prawdopodobieństwem p połączenie każdej z par wierzchołków [19]. Ze sposobu budowy grafu losowego wynika, że zawiera on przeciętnie

$$L = \langle |E| \rangle = p \binom{N}{2} = \frac{1}{2} pN(N-1) \quad (26)$$

krawędzi, w związku z tym średni stopień wierzchołka grafu będzie równy:

$$\langle k \rangle = \frac{2L}{N} = p(N-1) \approx pN. \quad (27)$$

Prawdopodobieństwo, że wybrany wierzchołek będzie połączony z k innymi spośród $N - 1$ wierzchołków jest równe prawdopodobieństwu uzyskania k sukcesów w $N - 1$ próbach niezależnych z prawdopodobieństwem sukcesu p i prawdopodobieństwem porażki $1 - p$, zatem rozkład wierzchołków w grafach losowych jest rozkładem dwumianowym, stąd:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (28)$$

Gdy $p \rightarrow 0$, rozkład dwumianowy (28) można przybliżyć rozkładem Poissona:

$$P(k) = \frac{e^{-\langle k \rangle} \langle k \rangle^k}{k!} \quad (29)$$

gdzie: symbol $\langle k \rangle$ oznacza średni stopień wierzchołka – patrz zależność (27).

Istotną wadą grafów losowych jest wąski rozkład stopni wierzchołków, wynikający z małej wartości wariancji, która dla rozkładu Poissona jest równa $\sigma_k^2 = k$, stąd odchylenie standardowe stopnia wierzchołka $\sigma_k = \sqrt{k}$. W konsekwencji wierzchołki o stopniach znacznie przekraczających średni stopień $\langle k \rangle$ w klasycznych grafach losowych nie pojawiają się. W związku z tym klasyczne grafy przypadkowe można traktować jak grafy regularne, w których wszystkie węzły mają ten sam stopień $\langle k \rangle$ [9]. Inną wadą grafów losowych, która uniemożliwia ich wykorzystanie w modelowaniu rzeczywistych sieci złożonych, jest bardzo mała wartość współczynnika gronowania (8). Z uwagi na zakładaną w procedurze budowy sieci niezależność krawędzi, współczynnik gronowania dla grafu - C , a także dla każdego węzła sieci - C_i jest równy prawdopodobieństwu utworzenia pojedynczej krawędzi, tj.

$$C = C_i = p. \quad (30)$$

W grafach losowych każde powstanie połączenia dwóch węzłów krawędzią można traktować jako zdarzenie niezależne, a zatem prawdopodobieństwo powstania krawędzi jest takie samo dla dowolnej pary węzłów. Wykazano, że klasyczne grafy przypadkowe są małymi światami. Przeciętna odległość l między dowolną parą węzłów w tych sieciach spełnia relację [9]:

$$l \sim \frac{\ln N}{\ln \langle k \rangle} \quad (31)$$

Innym modelem konfiguracyjnym grafów losowych są grafy losowe o zadanym rozkładzie stopni wierzchołków. Procedura tworzenia takich grafów obejmuje następujące kroki:

- ustalenie liczby wierzchołków N ;
- dołączenie do każdego wierzchołka $i = 1, 2, \dots, n$ połówek krawędzi, których liczba k_i jest losowana zgodnie z ustalonym rozkładem liczby krawędzi $P(k)$.
- łączenie połówek krawędzi w sposób losowy.

Efektom losowego łączenia krawędzi są różne realizacje sieci z zadanej na początku procedury sekwencją stopni węzłów $\{k_1, k_2, \dots, k_n\}$. Strukturalne własności tak utworzonych grafów losowych są wielkościami uśrednionymi po wszystkich mikroskopowych realizacjach tych grafów [9]. Ich podstawową charakterystyką jest średni stopień wierzchołka grafu:

$$\langle k \rangle = \sum_k k P(k) \quad (32)$$

oraz rozkład $Q(k)$ stopni wierzchołków, umieszczonych na końcach losowo wybranych krawędzi, określony na $N \cdot \langle k \rangle = 2|E|$ – elementowym zbiorze połówek krawędzi, dołączonych do wszystkich węzłów sieci. Rozkład $Q(k)$ opisuje prawdopodobieństwo dołączenia losowo wybranej połowki krawędzi do węzła o stopniu k , może być też traktowany jako rozkład stopni węzłów należących do najbliższego sąsiedztwa dowolnego wierzchołka (nie zależy od stopnia węzła, którego otoczenie jest analizowane). Określa go zależność:

$$Q(k) = \frac{k}{\langle k \rangle} P(k). \quad (33)$$

Wszystkie wierzchołki mają, statystycznie rzecz biorąc, takie same sąsiedztwo - usieciwienie najbliższych sąsiadów nie zależy od stopnia wierzchołka – nie ma znaczenia, czy badana się otoczenie huba, czy wierzchołka o stopniu $k = 1$ [9]. Średni stopień wierzchołka najbliższego sąsiada każdego z tych wierzchołków jest taki sam i wynosi:

$$\langle k \rangle_{nn} = \sum_k k Q(k) = \frac{\langle k^2 \rangle}{\langle k \rangle}. \quad (34)$$

Wynika stąd, że w sieciach o potęgowych rozkładach stopni wierzchołków, w których $\alpha \in (2, 3)$, średni stopień wierzchołka najbliższego sąsiada dowolnego węzła $\langle k_{nn} \rangle$ jest większy niż średni stopień wierzchołka, $\langle k \rangle$, dla całej sieci.

Niech p_{ij} oznacza prawdopodobieństwo istnienia krawędzi między parą węzłów i oraz j , o stopniach odpowiednio równych k_i i k_j . Jest ono równoważne prawdopodobieństwu, że w trakcie tworzenia sieci przynajmniej jedna spośród k_i połówek krawędzi dołączonych do węzła i połączy się z dowolną spośród k_j połówek krawędzi dołączonych do j , tj. [9]:

$$p_{ij} = 1 - \left(1 - \frac{k_j}{\langle k \rangle N}\right)^{k_i} \approx \frac{k_i k_j}{\langle k \rangle N}. \quad (35)$$

Znajomość rozkładu stopni wierzchołków w sąsiedztwie wybranego węzła $Q(k)$ oraz prawdopodobieństwa istnienia krawędzi p_{ij} umożliwia wyznaczenie parametrów strukturalnych sieci. Przykładowo, współczynnik gronowania dowolnego węzła w tych sieciach jest równy uśrednionemu, po najbliższym sąsiedztwie tego węzła, prawdopodobieństwu p_{ij} . W obliczeniach uwzględnia się fakt, że obydwa węzły i oraz j wykorzystają już po jednej ze swoich krawędzi, tworząc bezpośrednie połączenia do tego węzła, dla którego współczynnik gronowania ma być wyznaczony, stąd:

$$C = \sum_{k_i} \sum_{k_j} \frac{(k_i - 1)(k_j - 1)}{\langle k \rangle N} Q(k_i) Q(k_j) = \frac{1}{N} \frac{(\langle k^2 \rangle - k^2)}{\langle k \rangle^3} \quad (36)$$

Ponadto średnia odległość między dowolną parą węzłów w badanych sieciach losowych będzie równa [1]:

$$l = \frac{\ln(\langle k^2 \rangle - \langle k \rangle) - 2\langle \ln k \rangle + \ln N - \gamma}{\ln(\langle k^2 \rangle / \langle k \rangle) - 1} + \frac{1}{2}, \quad (37)$$

gdzie: $\gamma = 0,5772\dots$ – stała Eulera.

4. MODELE SIECI EWOLUJĄCYCH

W 1999 roku na Uniwersytecie Notre Dame w Indianie przeprowadzono badania, których efektem było odkrycie potęgowych rozkładów prawdopodobieństwa, opisujących liczby połączeń (linków) wychodzących ze stron WWW, a także wskazujących na strony tej sieci. Okazało się, że sieć WWW jest bezskalowa, w związku z tym opisując ją należy unikać określeń typu „wartości średnie”. Sieć WWW nie rozwija się według ściśle narzuconych reguł, a struktura występujących w niej połączeń jest efektem indywidualnych decyzji użytkowników, określających liczbę i rodzaj odnośników na tworzonej stronie WWW.

Okazało się, że inne sieci rzeczywiste (ang. real world networks) również wykazują potęgowy rozkład stopni wierzchołków - m.in. taka sytuacja występuje w przypadku Internetu, sieci telefonii komórkowej, wielu sieci społecznych, m.in. sieci współpracy naukowej, sieci cytowań, a nawet sieci kontaktów seksualnych [9]. Rozważania mające na celu poznanie odpowiedzi na zasadnicze pytanie, odnośnie przyczyny bezskalowości sieci złożonych przedstawili A.-L. Barabási i R. Albert w pracy pt. *Emergence of scaling in random networks* [20], w której wykazali, że potęgowy rozkład stopni wierzchołków w sieciach rzeczywistych jest spowodowany dwoma komplementarnymi mechanizmami: wzrostem sieci i regułami preferencyjnego dołączania węzłów [1]. Procedura

powstawania ewoluującej sieci Barabásiego-Albert składa się z następujących opisanych niżej etapów [9] – rys. 7:

- w początkowej chwili $t = 0$ sieć składa się z całkowicie połączonego klastra wierzchołków o rozmiarze $m_0 \geq 1$;
- w kolejnych krokach czasowych $t = 1, 2, 3, \dots$ do sieci dodawane są nowe wierzchołki (po jednym w każdym kroku), tworzące $m \leq m_0$ połączeń z istniejącymi wierzchołkami sieci, przy czym nowe wierzchołki są dołączane do już istniejących, zgodnie z liniową regułą preferencyjnego dołączania, polegającą na tym, że prawdopodobieństwo utworzenia połączenia z jednym ze starszych wierzchołków jest proporcjonalne do stopnia tego wierzchołka. Na przykład dla wierzchołka sieci o numerze i ze stopniem wierzchołka k_i :

$$\Pi(k_i) = \frac{k_i}{\sum_{j=1}^t k_j} = \frac{k_i}{2mt}; \quad (38)$$

- ewolucja sieci kończy się w pewnej chwili t .

W celu wyznaczenia rozkładu stopni wierzchołków w sieci Barabásiego-Albert (BA) wykorzystuje się tzw. metodę czasu ciągłego, w której oprócz założenia o ciągłości czasu dodatkowo przyjmuje się, że stopnie $k_i(t)$ wierzchołków sieci również zmieniają się w sposób ciągły, w związku z tym w opisywanej metodzie wykorzystuje się średnie wartości stopni poszczególnych wierzchołków, a nie ich dokładne wartości. Na podstawie tych średnich stopni wyznacza się rozkład stopni wierzchołków, z czego wynika, że otrzymany rozkład jest rozkładem uśrednionym. Ścisłą metodę wyznaczenia rozkładu stopni wierzchołków w sieci BA, można znaleźć m.in. w [7].

Uśredniony rozkład stopni wierzchołków opisuje zależność:

$$P(k) = \frac{2m^2}{k^3}. \quad (39)$$

Uzyskany rozkład jest rozkładem potęgowym o wykładniku charakterystycznym $\alpha = 3$. Proste modyfikacje założeń modelu BA prowadzą do rozkładów potęgowych o innej wartości wykładnika charakterystycznego [7]. Wyrażenie opisujące rozkład stopni wierzchołków w sieci BA ma postać:

$$P(k) = \frac{2m(m+1)}{k(k+1)(k+2)}, \quad (40)$$

gdzie $k \geq m$.

W granicy, przy $k \gg m$ rozkład (40), uzyskany metodą ścisłą, staje się równoważny rozkładowi opisanemu zależnością (39).

Modyfikacją ewoluującej sieci Barabásiego-Albert było uwzględnienie w regule preferencyjnego dołączania (38) parametru opisującego początkową atrakcyjność węzłów a , zgodnie z zależnością:

$$\Pi(k_i) = \frac{k_i + a}{\sum_{j=1}^t (k_j + a)} \quad (41)$$

Zastosowanie reguły (41) prowadzi do uzyskania sieci bezskalowych, z wykładnikiem charakterystycznym α , opisującym rozkład stopni wierzchołków, będącym funkcją parametru a :

$$\alpha = 3 + \frac{a}{m}. \quad (42)$$

Prawdopodobieństwo $\Pi(k_i)$ nie może być ujemne, co oznacza, że parametr $a > -m$, przy czym, analogicznie, jak w sieci BA, $m = k_i(t_i) \leq k_i(t_i)$ jest najmniejszą liczbą połączeń dołączonych do wierzchołka i . Po wstawieniu granicznych wartości parametru a do wyrażenia (42) otrzymuje się graniczne wartości wykładnika charakterystycznego α dla rozkładu stopni wierzchołków w badanych sieciach: $\alpha \in (2, \infty)$. Wynika stąd, że uzupełnienie liniowej reguły preferencyjnego dołączania węzłów w procedurze konstrukcyjnej sieci BA o początkową atrakcyjność węzłów umożliwi uzyskanie sieci bezskalowych o wykładniku charakterystycznym $\alpha \neq 3$ [9].

Najbardziej naturalnym uogólnieniem modelu BA było zastąpienie liniowej reguły preferencyjnego dołączania węzłów regułą nieliniową o postaci:

$$\Pi(k_i) = \frac{k_i^\gamma}{\sum_{j=1}^t k_j^\gamma} \quad (43)$$

przy czym $\gamma > 0$. Przyjęcie powyższej reguły znacząco wpływa na charakter rozkładu stopni wierzchołków sieci. Sieci bezskalowe uzyskuje się tylko dla $\gamma = 1$ (przypadek sieci BA). Dla $\gamma < 1$ uzyskuje się rozkład potęgowy z obcięciem wykładniczym, a dla $\gamma > 1$ można zaobserwować zjawisko kondensacji, polegające na pojawieniu się w tych sieciach pojedynczego superwęzła, skupiającego makroskopową liczbę krawędzi układu. Dla $\gamma > 2$ istnieje skończone prawdopodobieństwo, że taki wierzchołek (ang. *gel-node*) będzie połączony ze wszystkimi węzłami sieci [9].

W opisanym wyżej modelu BA i w jego uogólnieniach, dodawane do sieci nowe wierzchołki miały jednakowe własności - jednakową początkową liczbę krawędzi, a nawet jednakową początkową atrakcyjność. W układach rzeczywistych prawie nigdy nie mamy do czynienia z taką sytuacją. Nowe węzły dodawane do sieci rzeczywistych zawsze cechuje pewna różnorodność adaptacyjna (ang. *fitness*). Istnieje możliwość takiej modyfikacji reguły preferencyjnego dołączania węzłów w sieciach BA, by odzwierciedlała ona wspomnianą różnorodność adaptacyjną wierzchołków sieci rzeczywistych:

$$\Pi(k_i, \eta_i) = \frac{\eta_i k_i}{\sum_{j=1}^t \eta_j k_j} \quad (44)$$

W zależności (44) η_i jest parametrem adaptacyjnym – czyli zmienną losową o rozkładzie $\rho(\eta)$, przypisywanym węzłowi w chwili dodawania go do sieci. Usieciowienie $k_i(t)$ wierzchołka i nie zależy teraz wyłącznie od momentu t , w którym został dodany do sieci, bowiem nowe, lepiej przystosowane wierzchołki o dużych wartościach parametru η_i mogą znacznie szybciej niż stare gromadzić nowe połączenia. W pracy [7] pokazano, że jednorodny rozkład parametrów dopasowania $\rho(\eta) = \text{const}$ prowadzi do rozkładów stopni wierzchołków mających tłuste ogony, a ściślej do rozkładów potęgowych z logarytmiczną korektą:

$$P(k) \propto \frac{k^{-1+C}}{\ln k} \quad (45)$$

ze stałą $C = 1,255$ wyznaczoną numerycznie [9].

Poważnym ograniczeniem w opisywanych wyżej sieciach ewoluujących i ich modyfikacjach jest fakt, że liczba krawędzi $|E|$ była liniową funkcją liczby węzłów N , w związku z czym średni stopień wierzchołka $\langle k \rangle$ nie zależy od czasu. W rzeczywistych sieciach złożonych występują inne własności – np. liczba krawędzi przyrasta w nich szybciej niż liczba wierzchołków. Takie sieci określa się jako sieci o przyspieszonym wzroście. W większości wypadków przyspieszony wzrost sieci może być związany z innymi czynnikami – mogą np. powstawać nowe połączenia pomiędzy starymi wierzchołkami sieci, a inne krawędzie, reprezentujące relacje między wierzchołkami, mogą zniknąć. W innych wariantach sieci z przyspieszonym wzrostem liczba nowych połączeń m , jakie tworzą wierzchołki w chwili dodania do sieci, mogą być potęgowo zależne od jej rozmiaru: $m(t) \propto t^\alpha$, gdzie $\alpha > 0$. W przypadku przyjęcia takiej reguły preferencyjnego dołączania, opisanej zależnością:

$$\Pi(k, t) \propto k + Bt^\alpha \quad (46)$$

sieci mają potęgowy rozkład stopni wierzchołków z wykładnikiem charakterystycznym $\alpha = 2 + B(1 + a)/(1 - Ba)$ [9].

Opisane wyżej reguły preferencyjnego dołączania mają zasadniczą wadę - dołączanie nowych wierzchołków wynika ze znajomości stopni wszystkich wcześniej dodanych węzłów, a z reguły taka informacja nie jest dostępna dla pojedynczego węzła. W związku z tym opracowano modele sieci ewoluujących, w których lokalne mechanizmy działają analogicznie do zasady preferencyjnego dodawania węzłów, przy czym nie wymaga to znajomości struktury całego układu. To właśnie

lokalne struktury wzrostu w tych sieciach narzucają bezskalowy charakter rozkładu stopni wierzchołków.

Model sieci skierowanych z mechanizmem kopiowania opisuje ewolucję sieci WWW, w których tworzenie nowych stron WWW zwykle ma związek z częściowym kopiowaniem już istniejących stron i znajdujących się na nich hyperlinków. Procedura konstrukcyjna z mechanizmem kopiowania obejmuje następujące kroki [9]:

- wzrost sieci rozpoczyna się w chwili $t = 0$ od pierwotnego klastra, zbudowanego z $m_0 \geq 1$ wierzchołków, przy czym każdy z nich posiada $m < m_0$ krawędzi, łączących je z innymi, losowo wybranymi wierzchołkami tego klastra;
- w kolejnych krokach czasowych $t = 1, 2, 3, \dots$ dołączane są nowe wierzchołki (w każdym kroku jeden), w taki sposób, że w chwili powstania każdy wierzchołek ma $k^{in} = 0$ połączeń wchodzących i $k^{out} = m$ połączeń wychodzących, z docelowymi wierzchołkami kopiowanymi z losowo wybranego, starszego wierzchołka tej sieci; końce wychodzących krawędzi z prawdopodobieństwem p dołączane są do innych, losowo wybranych wierzchołków;
- wzrost sieci kończy się w chwili t .

W tak zdefiniowanym modelu każdy wierzchołek sieci ma dokładnie m wychodzących połączeń:

$$P_{out}(k^{out}) = \delta(k^{out} - m), \quad (47)$$

a rozkład liczby połączeń wchodzących do dowolnego wierzchołka jest opisany prawem potęgowym:

$$P_{in}(k^{in}) \propto (k^{in})^{-\alpha}, \quad (48)$$

z wykładnikiem charakterystycznym

$$\alpha = \frac{2-p}{1-p} \quad (49)$$

należącym do przedziału $(2, \infty)$.

Równanie, opisujące tempo zmiany w czasie średniego stopnia dowolnego węzła, ma postać:

$$\frac{\partial k^{in}}{\partial t} = (1-p) \frac{k^{in}}{t} + mp \frac{1}{t}. \quad (50)$$

W powyższym równaniu k^{in} oznacza wejściowy stopień wierzchołka, dodanego do sieci w chwili t_i . Stopień ten wzrasta o 1, gdy wierzchołek jest sąsiadem węzła, którego krawędzie zostały w danym kroku czasowym skopiowane, z prawdopodobieństwem równym k_{in}/t , przy czym skopiowane połączenie do tego węzła nie zostało przełączone (prawdopodobieństwo tego zdarzenia jest równe $1-p$). Drugi wyraz po prawej stronie równania (50) opisuje średnią liczbę tych spośród przełączonych krawędzi, które trafiły do rozważanego węzła [9].

Rozwiązanie równania różniczkowego (50) przy warunku początkowym $k_{in}(t_i) = 0$ jest następujące:

$$k^{in}(t) = \frac{mp}{(1-p)} \left[\left(\frac{t}{t_i} \right)^{1-p} - 1 \right]. \quad (51)$$

Opisuje ono zmienność stopnia wejściowego dowolnego wierzchołka sieci, gdy ta sieć rośnie. Na podstawie tej zależności można wyznaczyć rozkład stopni wierzchołków $P_{in}(k_{in})$.

W pracy [1] pokazano, że lokalny mechanizm kopiowania jest równoważny regule preferencyjnego dołączania węzłów, stanowiącej podstawę modelu BA. Losowo wybrane krawędzie są częściej dołączane do wierzchołków wyższych stopni – kopiowanie przez nowy wierzchołek połączeń wychodzących z losowo wybranych starszych wierzchołków przynosi efekt podobny do dołączania tych krawędzi do preferencyjnie wybranych wierzchołków.

W sieciach Barabásiego-Albert prawdopodobieństwo przyłączenia nowej krawędzi zależy od stopnia docelowego wierzchołka, w konsekwencji starsze wierzchołki mają z czasem coraz wyższy stopień i nigdy się nie starzeją. W sieciach rzeczywistych, a szczególnie w sieciach społecznych (np. sieci współpracy naukowej, sieci cytowań) zdolność starszych węzłów do przyjmowania nowych krawędzi maleje z czasem, a nawet może dojść do sytuacji, w której węzeł staje się nieaktywny.

Poniżej przedstawiono procedurę tworzenia modelu sieci ewoluującej z uwzględnieniem efektów starzenia się wierzchołków:

- w chwili $t = 0$ sieć obejmuje całkowicie połączony klaster o rozmiarze $m_0 \geq 1$, ze wszystkimi wierzchołkami aktywnymi;
- w kolejnych krokach czasowych $t = 1, 2, 3, \dots$ do sieci dodawane są nowe wierzchołki z m połączeniami z aktywnymi wierzchołkami tej sieci, przy czym w każdym kroku czasowym w sieci jest jedynie m aktywnych węzłów; dodanie do sieci nowego wierzchołka jest równoznaczne z dezaktywacją jednego z $m + 1$ aktywnych wierzchołków; jednocześnie zakłada się, że prawdopodobieństwo dezaktywacji wierzchołka i jest funkcją jego stopnia k_i :

$$q(k_i) = \frac{\gamma - 1}{a + k_i}, \quad (52)$$

przy czym $\gamma - 1 = \left[\sum_{i=1}^t (a + k_i)^{-1} \right]^{-1}$ - jest stałą normalizacyjną, zaś $a > -m$ jest odpowiednio dobraną stałą;

- zakończenie wzrostu sieci następuje w pewnej chwili t .
Efektem działania powyższej procedury są sieci bezskalowe o rozkładzie stopni wierzchołków opisanym zależnością $P(k) \propto k^{-\alpha}$ z wykładnikiem charakterystycznym $\alpha > 2$ i wysokim współczynnikiem gronowania $C = 5/6$ [7]. W modelu z efektem starzenia się sieci zasada preferencyjnego dołączania wierzchołków, skutkująca bezskalowym charakterem tworzonych sieci, jest konsekwencją lokalnych reguł wzrostu tych sieci, bowiem nowo dodany wierzchołek posiada tylko cząstkową informację o strukturze sieci, ograniczoną do małej grupy aktywnych wierzchołków stanowiących lokalne środowisko tego węzła.

Ostatnią omawianą w niniejszej pracy klasą modeli ewoluujących są ważne sieci bezskalowe. Procedura tworzenia takich sieci składa się z następujących etapów [7]:

- rozwój sieci rozpoczyna się od całkowicie połączonych klastera, złożonego z $m_0 \geq 1$ wierzchołków z początkowymi wagami wszystkich krawędzi równymi w_0 ;
- w każdym kolejnym kroku czasowym $t = 1, 2, 3, \dots$ do sieci dodawany jest jeden nowy wierzchołek; tworzą one $m \leq m_0$ nowych połączeń z wcześniej utworzonymi wierzchołkami tej sieci; nowe wierzchołki tworzone są zgodnie z liniową regułą preferencyjnego dołączania, przy uwzględnieniu siły docelowego węzła s_i , a nie jego stopnia k_i , zgodnie z zależnością:

$$\Pi(s_i) \propto s_i; \quad (53)$$

- nowym krawędziom przypisywana jest taka sama waga w_0 , skutkiem czego w chwili dodania do sieci, siła nowego wierzchołka jest równa $s_i(t_i) = mw_0$ przy stopniu wierzchołka równym $k_i(t_i) = m$; ponadto dołączenie do starego wierzchołka nowej krawędzi, poza zmianą stopnia i siły tego wierzchołka odpowiednio równymi $\Delta k_i = 1$ oraz $\Delta s_i = w_0$, skutkuje zmianą siły wierzchołka o dodatkową wartość δ reprezentującą dodatkowy wzrost siły węzła spowodowany dołączeniem do niego nowej krawędzi:

$$s_i(t+1) = s_i(t) + w_0 + \delta; \quad (54)$$

- przyjmuje się ponadto reguły zmiany wag krawędzi: jeśli do wierzchołka i została dołączona nowa krawędź, to wagi pozostałych krawędzi tego wierzchołka ulegają modyfikacji zgodnie z zależnością:

$$w_{ij}(t+1) = w_{ij}(t) + \Delta w_{ij}, \quad (55)$$

przy czym zakłada się, że zmiana Δw_{ij} jest proporcjonalna do wyjściowej wagi połączeń, tj.

$$\Delta w_{ij} = \delta \frac{w_{ij}(t)}{\sum_j w_{ij}(t)} = \delta \frac{w_{ij}(t)}{s_i(t)}; \quad (56)$$

- koniec ewolucji sieci może nastąpić w dowolnej chwili t .

Aby wyznaczyć rozkład sił wierzchołków w ważonych sieciach bezskalowych wykorzystuje się wspomnianą wcześniej metodę czasu ciągłego. W wyniku jej zastosowania otrzymuje się potęgowy rozkład sił wierzchołków $P(s) \propto s^{-\alpha}$, przy czym:

$$\alpha = \frac{4\delta + 3}{2\delta + 1}. \quad (57)$$

Dla nieujemnych wartości δ wykładnik rozkładu (57) zmienia się w granicach od $\alpha = 2$ do $\alpha = 3$, co jest charakterystyczne dla potęgowych rozkładów sił wierzchołków w sieciach rzeczywistych.

Korzystając z metody czasu ciągłego można wykazać, że rozkład potęgowy opisuje również rozkład stopni wierzchołków oraz rozkład wag połączeń międzywęzłowych. Takie własności są charakterystyczne dla wielu rzeczywistych sieci ważonych, a w szczególności sieci połączeń lotniczych [7].

WNIOSKI

Termin „sieci złożone” pochodzi od S.N. Dorogovtseva i J.F.F. Mendesa, autorów książki „Evolution of Networks. From Biological Nets to the Internet and WWW” [7]. W literaturze pojęcie to pojawiło się pod koniec lat 90. ubiegłego wieku, gdy badacze, zajmujący się różnymi dziedzinami nauki – informatycy, biolodzy, socjolodzy, fizycy, matematycy – zaczęli opisywać i modelować różne rzeczywiste sieci, z którymi stykali się w swej działalności. Termin ten odnosi się do sieci, charakteryzujących się bardziej skomplikowaną architekturą, niż sieci opisywane za pomocą grafów losowych z określoną (zadaną z góry) liczbą wierzchołków i połączeń (krawędzi). Charakterystyczną cechą sieci złożonych jest obecność w nich silnie połączonego zespołu węzłów (tzw. hubu), odgrywającego w tych sieciach kluczową rolę. Przyjęcie takiego opisu topologii sieci sprawia, że znakomitą większość współcześnie funkcjonujących sieci można uznać za sieci złożone. Zainteresowanie badaczy tymi sieciami nie wynika wyłącznie z faktu zainteresowania sieciami złożonymi o kluczowym obecnie znaczeniu – Internetem i sieciami telekomunikacyjnymi. Najważniejszym powodem zainteresowania sieciami złożonymi jest fakt, że ich własności topologiczne i struktura różnią się od własności i struktury opisanych już i dość dobrze zbadanych sieci o strukturze kraty – sieci i ich funkcjonowanie nie mogą być analizowane w oparciu o teorie opracowane dla krat o skończonej liczbie wymiarów – konieczne jest, zatem, nowe spojrzenie na analizę sieci złożonych i opracowanie nowych metod badawczych, opartych przede wszystkim na metodach fizyki i mechaniki statystycznej.

Przedstawione wyżej charakterystyki i własności sieci złożonych mogą być wykorzystane do efektywnego i wydajnego opisu sieci transportowych (np. sieci komunikacji publicznej na określonym obszarze, sieci dróg publicznych, sieci kolejowej itp.), jak również sieci logistycznych (np. sieci zaopatrzenia i dystrybucji dużych organizacji handlowych, sieci przesyłek pocztowych). Z tego opisu teoretycznego wynikają konkretne wnioski o charakterze praktycznym – np. w sieci transportu publicznego można rozważać różne konfiguracje sieci w celu identyfikacji jej słabych miejsc (wąskich gardeł), poszukiwania rozwiązań alternatywnych, zapewniających jej lepsze funkcjonowanie. W Politechnice Świętokrzyskiej prowadzone są obecnie badania ukierunkowane na budowę modeli sieci transportowych, oparte na przedstawionych wyżej rozważaniach teoretycznych.

Streszczenie

Praca poświęcona jest charakterystyce problematyki modelowania sieci złożonych, mogących znaleźć zastosowanie przy opisie i badaniach różnego rodzaju systemów funkcjonujących we współczesnym świecie, w tym systemów transportowych i logistycznych. Przedstawiono wyniki badań ukierunkowanych na modele sieci ewoluujących, które stanowią aktualnie najlepsze przybliżenie opisu sieci rzeczywistych. Scharakteryzowano różne modele sieciowe, poczynając od sieci, bazujących na grafach losowych, a kończąc na stosunkowo niedawno opisanych modelach sieci ewoluujących. Zdaniem autorów opisane w pracy modele sieci złożonych mogą być praktycznie wykorzystane do opisu sieci transportowych i logistycznych, a badanie modeli takich sieci może przynieść wymierne korzyści praktyczne.

Complex networks as the transportation and logistics networks modelling tools

Abstract

The aim of the contribution was to characterize a problem of complex networks modeling, which can be used for description and a study of various types of systems functioning in the modern world, including transport and logistical systems. The results of research concentrated on evolving network models, which are currently the best approximation of actual description of the network have been presented. Various network models, starting from the network models, basing on random graphs, and ending with the relatively recently described models of evolving networks have been described. According to the authors the models of complex networks presented in the contribution can be practically used for the transport and logistical systems description, and test models of such networks can bring tangible practical benefits.

BIBLIOGRAFIA

1. Frączak A., Frączak P.: *Świat sieci złożonych. Od fizyki do Internetu*. Wydawnictwo Naukowe PWN, (2009).
2. Alexanderson, G.: *Euler and Königsberg's bridges: a historical view*. Bulletin of the American Mathematical Society **43** (4): 567 (2006).
3. Sylvester J. J.: *Chemistry and algebra*, Nature, **17**: 284 (1878).
4. Moreno, J. L., *Who Shall Survive?*, New York, N.Y.: Beacon House (1934).
5. Erdős, P. Rényi, A. *On Random Graphs I*, in: Publ. Math. Debrecen 6, p. 290–297 (1959).
6. Watts, D.J.; Strogatz, S.H.: *Collective dynamics of 'small-world' networks*. Nature **393** (6684): 440–442 (1998).
7. Dorogovtsev S.N., Mendes J.F.F.: *Evolution of Networks. From Biological Nets to the Internet and WWW*. Clarendon Press, Oxford (2002).
8. *The size of the World Wide Web (The Internet)*: <http://www.worldwidewebsite.com/>
9. *Fizyka sieci złożonych*. <http://www.if.pw.edu.pl/~agatka/sieci.html>
10. Grimmett, G.: *Percolation*, Springer (1999)
11. Newman M.E.J.: *Who is the best connected scientist? A study of scientific coauthorship networks, Scientific collaboration networks. I. Network construction and fundamental results*, Phys. Rev. E **64**, 016131; *II. Shortest paths, weighted networks, and centrality*, Phys. Rev. E **64**, 016131 (2001).
12. Newman, M.E.J.: *Networks: An Introduction*. Oxford University Press (2010).
13. Sabidussi, G.: *The centrality index of a graph*. Psychometrika **31**, 581–603 (1966).
14. Dangalchev Ch., *Residual Closeness in Networks*, Physica A **365**, 556 (2006).
15. *Elementy teorii grafów. Statystyczne charakterystyki sieci*: <http://www.if.pw.edu.pl/~agatka/moodle/charakterystyki.html> (5.02.2014).
16. *How Google Finds Your Needle in the Web's Haystack*: <http://www.ams.org/samplings/feature-column/fcarc-pagerank> (5.02.2014).
17. Newman M. E. J.: *The mathematics of networks*. <http://www-personal.umich.edu/~mejn/papers/palgrave.pdf> (5.02.2014).
18. Mańka-Krasoń A.: *Wpływ frustracji na własności sieci Erdősa-Rényi z oddziaływaniem antyferromagnetycznym w modelu Isinga*. Rozprawa doktorska, Akademia Górniczo-Hutnicza, Wydział Fizyki i Informatyki Stosowanej, Kraków (2010).
19. van den Hofstad, R.: *Random Graphs and Complex Networks. Lecture Notes*. Eindhoven University of Technology, Eindhoven (2013) <http://www.win.tue.nl/~rhofstad/NotesRGCN.pdf>.
20. Barabási A.-L. and Albert R.: *Emergence of Scaling in Random Networks*. Science, **286**:509-512 (1999)
21. Newman M.E.J.: *The structure and function of complex networks*. <http://arxiv.org/pdf/condmat/0303516.pdf> (5.02.2014).