

Czy sieci rządzą światem? Od Eulera do Barabasiego

Z. TARAPATA

zbigniew.tarapata@wat.edu.pl

Instytut Systemów Informatycznych, Wydział Cybernetyki
Wojskowa Akademia Techniczna
ul. Gen. S. Kaliskiego 2, 00-908 Warszawa

W pracy poruszono problem różnorodnych zastosowań modeli i metod teorii grafów i sieci. W szczególności zwrócono uwagę na cechy charakterystyczne współczesnych sieci rzeczywistych. Przeanalizowano, w jaki sposób modeluje się te sieci, jakie mają własności i co z tego wynika. Poruszono algorytmiczne aspekty sieci złożonych ukierunkowane na złożoność obliczeniową metod wyznaczania wartości charakterystyk sieci. Kilkanaście interdyscyplinarnych przykładów rzeczywistych sieci zilustrowało opisywane własności sieci. We wnioskach opisano tendencje rozwojowe współczesnej teorii grafów i sieci, zwłaszcza w kontekście modelowania ewolucji sieci złożonych.

Słowa kluczowe: sieci złożone, systemy złożone, bezskalowość, małe światy, zastosowania sieci złożonych, algorytmiczne aspekty sieci złożonych.

1. Wprowadzenie

Pytanie tytułowe, choć postawione w sposób dość filozoficzny, jak okaże się w dalszej części artykułu, z filozofią ma niewiele wspólnego, za to bardzo dużo z jednym z działów matematyki dyskretniej oraz obserwacją świata rzeczywistego. Autor, podpierając się szeregiem różnych przykładów praktycznych oraz wynikami najnowszych obserwacji i badań, postara się zasugerować Czytelnikowi odpowiedź na to pytanie.

Jednym z działów matematyki dyskretniej jest teoria grafów i sieci, która bardzo prędko się rozwija, a z racji swoich interdyscyplinarnych zastosowań znajduje szerokie uznanie w świecie nauki i praktyki [6], [8].

Podstawowe problemy teorii grafów i sieci oraz ich zastosowania opisano poniżej:

- problemy kolorowania grafów: harmonogramowanie, układanie planów, szeregowanie zadań, kolorowanie map i inne
- problemy dróg ekstremalnych: planowanie podróży, wyszukiwanie najkrótszych tras, routing, sieci semantyczne i inne
- problemy analizy sieci PERT, GERT: analiza złożonych przedsięwzięć
- problemy przepływów: przepływy w sieciach (komputerowych, gazowych, wodociągowych, itp.) i inne
- problemy przydziałów: przydziały częstotliwości, pracowników do maszyn, kojarzenie małżeństw i inne

- problemy podobieństwa grafów: rozpoznawanie wzorców strukturalnych, wyszukiwanie synonimów w bazach danych (słownikowych), wyszukiwanie podobnych stron WWW, ontologie
- analiza sieci złożonych: Internet, sieci społeczne, biologiczne, komputerowe, transportowe, energetyczne i inne.

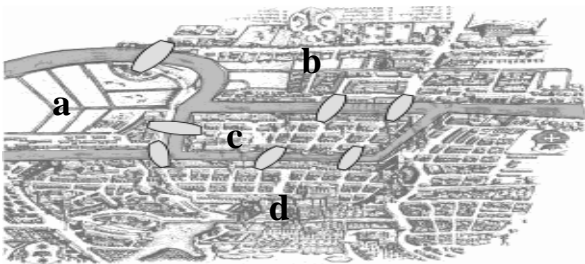
Interdyscyplinarność zastosowań modeli i metod teorii grafów i sieci (widoczna szczególnie w opisie ostatniej grupy zastosowań) skłania do bliższego przyjrzenia się tej tematyce oraz wypracowania odpowiedzi na pytanie tytułowe. Artykuł zorganizowany jest w następujący sposób. W rozdziale 2 krótko zaprezentowano genezę grafów. Rozdział 3 zawiera szereg przykładów współczesnych sieci, wraz z krótkim opisem sposobu ich powstawania. W rozdziale 4 opisano charakterystykę systemów złożonych, które modeluje się z wykorzystaniem sieci złożonych, o pewnych specyficznych własnościach. Rozdział 5 zawiera dalsze przykłady sieci rzeczywistych zaskakujące swą oryginalnością i dziedziną zastosowań. W rozdziale 6 opisano przykłady projektów realizowanych na Wydziale Cybernetyki WAT, w których wykorzystano modele sieciowe. Artykuł kończą wnioski oraz opis trendów rozwojowych teorii grafów.

2. Krótka historia teorii grafów

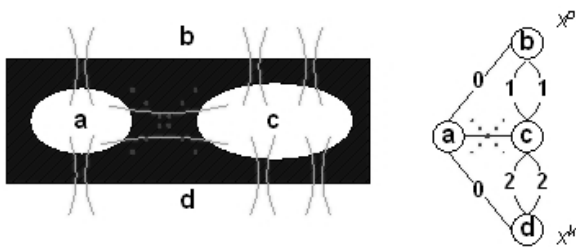
Za ojca teorii grafów uchodzi szwajcarski matematyk i fizyk Leonhard Euler (1707–1783), który w roku 1736 sformułował tzw. problem mostów królewieckich [11], używając przy tym pierwszy raz specyficznego modelu, który nazwany został grafem.

Rzeka Pregola przepływająca przez Königsberg (przedwojenny Królewiec, obecnie Kaliningrad) dzieliła miasto na cztery obszary (a, b, c, d), jak na rysunku 1. Jednocześnie istniało siedem mostów, które pozwalały przejść na drugą stronę rzeki z jednego z tych czterech obszarów.

Pytanie, które postawił Euler, było następujące: czy można zaplanować trasę przejścia przez miasto, wychodząc z któregoś obszaru, w taki sposób, aby przejść przez każdy most dokładnie jeden raz? [11], [35]. Euler zaproponował, aby przedstawić sytuację w Königsbergu z rysunku 1 za pomocą specyficznego modelu matematycznego, który nazwany został grafem (rysunek 2).



Rys. 1. Fragment mapy Königsberga z 1736 r. wraz z zaznaczonymi siedmioma mostami
Źródło: opracowanie własne na podstawie http://en.wikipedia.org/wiki/File:Konigsberg_bridges.png



Rys. 2. Uproszczony model sytuacji z rysunku 1 (po lewej) i jego reprezentacja grafowa (po prawej)
Źródło: opracowanie własne

Najogólniej rzecz ujmując, definiuje się dwa podstawowe rodzaje grafów [6], [8]: graf skierowany i nieskierowany.

Graf skierowany (ang. *directed graph*) G_s :

$$G_s = \langle V, A \rangle \quad (1)$$

$$A \subset V \times V$$

Graf nieskierowany (ang. *undirected graph*) G :

$$G = \langle V, E \rangle \quad (2)$$

$$E \subset \{\{x, y\} \subset V\}$$

gdzie:

V – zbiór wierzchołków grafu (ang. *set of nodes, vertices*),

E – zbiór krawędzi grafu (ang. *set of edges*),

A – zbiór łuków grafu (ang. *set of arcs*) – nazywanych czasami krawędziami skierowanymi.

Niech $|V| = N$, $|E| = M$ ($|A| = M$).

Zgodnie z powyższymi definicjami krawędź grafu definiowana jest przez podzbiór dwuelementowy zbioru wierzchołków (kolejność elementów w podzbiorze nieistotna – własność zbioru), natomiast łuk grafu przez parę uporządkowaną wierzchołków (kolejność elementów w parze istotna – własność pary uporządkowanej).

Euler do swoich rozważań wykorzystał graf nieskierowany, w którym wierzchołki reprezentowały owe cztery obszary miasta, a krawędzie (reprezentujące mosty) występowały między tymi parami wierzchołków (obszarów), które te mosty łączyły (rysunek 2).

Pytanie Eulera w języku teorii grafów można sformułować następująco: czy w grafie z rysunku 2 istnieje droga¹ lub cykl² Eulera³?

Euler sformułował problem i udowodnił tezę, że warunkiem koniecznym do istnienia w grafie cyklu Eulera jest, aby wszystkie wierzchołki w grafie miały parzysty stopień⁴ (dodatkowo, jeżeli istnieją dokładnie dwa wierzchołki o nieparzystym stopniu, wówczas w grafie istnieje droga Eulera, w takiej sytuacji

¹ Droga (ang. *path*) w grafie to ciąg naprzemienny wierzchołków i „stykających” się z nimi krawędzi (lub wychodzących z nich łuków), który rozpoczyna się w podanym wierzchołku początkowym x^p , a kończy w podanym wierzchołku końcowym x^k , przy czym krawędzie (łuki) nie mogą się powtarzać. Dodatkowo po łukach możemy się poruszać wyłącznie zgodnie z ich zwrotem.

² Cykl (ang. *cycle*) w grafie to taka droga, która rozpoczyna się i kończy w tym samym wierzchołku.

³ Droga (cykl) Eulera (ang. *eulerian path, cycle*) to taka droga (cykl), która zawiera wszystkie krawędzie (łuki) grafu.

⁴ Stopień wierzchołka (ang. *node degree*) jest to liczba wszystkich gałęzi (krawędzi i łuków), którymi dany wierzchołek połączony jest z innymi wierzchołkami (czasami mówi się o liczbie sąsiadów wierzchołka).

jeden z tych dwóch wierzchołków jest początkiem tej drogi, a drugi końcem). Nie podał jednakże algorytmu rozwiązania sformułowanego przez siebie problemu.

W 1873 r. ukazała się praca Carla Hierholzera [15], już po jego śmierci, w której podał on dowód na to, że każdy graf spójny, w którym każdy wierzchołek ma parzysty stopień – posiada cykl Eulera (sformułował i udowodnił zatem warunki wystarczające do istnienia cyklu Eulera). W pracy tej Hierholzer podał również pierwszy algorytm wyznaczania takiego cyklu (Euler algorytmu nie podał).

Wracając do problemu mostów królewieckich, z tezy Eulera wynika, że nie istnieje rozwiązanie tego problemu (wszystkie cztery wierzchołki mają nieparzysty stopień). Gdyby na rysunku 2 nie było mostu przekreślonego przerywaną linią, wówczas istniałaby możliwość zaplanowania trasy przejścia w taki sposób, aby każdy most odwiedzić dokładnie jeden raz (liczba wierzchołków grafu o nieparzystych stopniach byłaby równa 2, tzn. wierzchołki b, d; jeden z nich byłby wierzchołkiem początkowym drogi, a drugi końcowym). Droga ta mogłaby wyglądać następująco: $b - (b, a) - a - (a, d) - d - (d, c) - c - (c, b) - b - (b, c) - c - (c, d) - d$. Grafy eulerowskie⁵ wykorzystywane są m.in. w bioinformatyce do rekonstrukcji sekwencji DNA z jej fragmentów [24], w problemie chińskiego listonosza, w przemyśle do planowania ruchu (trasy) ramion automatów przemysłowych [16], itd.

Warto jeszcze tylko zaznaczyć różnicę formalną między pojęciami *graf* i *sieć*. Sieć definiuje się jako graf opisany ilościowo (ważony), tzn. taki, w którym na wierzchołkach i/lub na krawędziach/łukach opisano jakieś funkcje, które mają interpretację zależną od rodzaju obiektu modelowanego przez sieć (np. w sieci transportowej: czas przejazdu po odcinku drogowym modelowanym przez krawędź/łuk, długość tego odcinka, itp.). Często potocznie, choć z formalnego punktu widzenia niepoprawnie, używa się zamiennie pojęć graf i sieć. W dalszej części artykułu, tam gdzie nie będzie to prowadzić do nieporozumień, będziemy używać zamiennie pojęć graf i sieć, pamiętając jednak, że graf opisuje wyłącznie strukturę jakiegoś obiektu (systemu) rzeczywistego, a sieć oprócz struktury opisuje charakterystyki ilościowe tego obiektu (systemu).

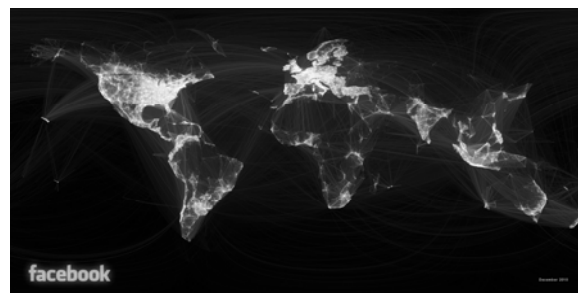
⁵ Graf eulerowski (ang. *Euler graph*) to taki, w którym istnieje cykl Eulera. Graf półeulerowski (ang. *semi-Euler graph*) to taki, w którym istnieje droga Eulera.

3. Przykłady współczesnych sieci rzeczywistych

W tym rozdziale zaprezentowanych zostanie kilkanaście przykładów sieci rzeczywistych z różnych dziedzin. Interdyscyplinarność tych zastosowań nie jest niczym nadzwyczajnym – jest to cecha współczesnego świata, praw, które nim rządzą, oraz charakteru zjawisk modelowanych za pomocą grafów i sieci.

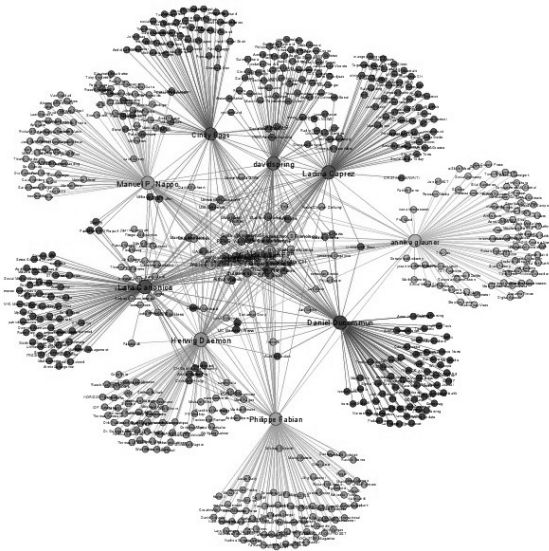
3.1. Internet

Pierwsza grupa przykładów dotyczy Internetu. Na rysunku 3 przedstawiono sieć geolokalizacji ok. 5 mln. użytkowników Facebooka (stan z grudnia 2010 r.). Każda geolokalizacja reprezentowana jest przez biały punkt na rysunku (jest to jednocześnie wierzchołek w grafie). Krawędziami w tym grafie są znajomości między dwoma użytkownikami Facebooka. Ciekawe jest to, że na rysunku nie umieszczono żadnego podkładu mapowego. Zarysy kontynentów, a w niektórych przypadkach i państw (np. USA), które są wyraźnie widoczne, powstały samoistnie z „chmury” białych punktów.



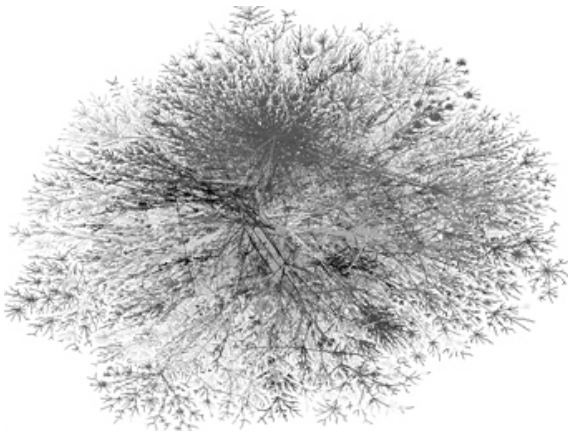
Rys. 3. Sieć geolokalizacji ok. 5 mln. użytkowników Facebooka z grudnia 2010 r.; Źródło: [36]

Na rysunku 4 przedstawiono sieć wybranych użytkowników Twittera wraz z ich zwolennikami (tzw. *followers*). Wierzchołkami w tej sieci są użytkownicy, a krawędziami połączono pary użytkownik-zwolennik (*follower*). Na tym przykładzie bardzo jaskrawo widać charakterystyczną cechę tej sieci: stosunkowo mało użytkowników ma bardzo dużo zwolenników, a zdecydowana większość użytkowników ma mało zwolenników. Cecha ta została opisana szczegółowo w rozdziale 4.4.

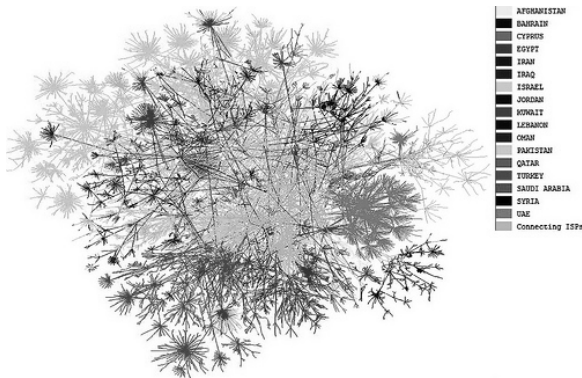


Rys. 4. Sieć wybranych użytkowników Twittera wraz z ich zwolennikami (tzw. *followers*); Źródło: [37]

Rysunki 5 i 6 zawierają fragment mapy ruchu w Internecie (rysunek 6 dotyczy krajów Środkowego Wschodu). Wierzchołki oznaczają komputery (routery), a krawędzie łączą te pary wierzchołków, między którymi odbywał się ruch (przekazywano informacje, pakiety danych itp.).



Rys. 5. Fragment mapy ruchu w Internecie; Źródło: [38]



Rys. 6. Mapa Internetu dla krajów Środkowego Wschodu; Źródło: [50]

3.2. Transport

Grupa przykładów w tym rozdziale dotyczy transportu. Na rysunku 7 przedstawiono sieć metra w Paryżu. Wierzchołkami w tej sieci są stacje metra, a krawędzie łączą stacje sąsiadujące geograficznie ze sobą, które znajdują się na trasie określonej linii metra.



Rys. 7. Sieć metra w Paryżu; Źródło: [39]

Rysunek 8 przedstawia fragment sieci ulicznej w São Paulo. Wierzchołkami tej sieci są skrzyżowania, a krawędzie (bądź łuki) łączą dwa skrzyżowania, między którymi istnieje odcinek drogowy. Istotne jest, że w przypadku sieci drogowej jako model połączenia występują najczęściej łuki, dzięki którym można zamodelować odcinki drogowe jednokierunkowe.



Rys. 8. Sieć uliczna w São Paulo; Źródło: [40]

Na rysunku 9 przedstawiono sieć szybkich kolei w Europie (stan na rok 2011). Wierzchołkami tej sieci są ważne (największe) węzły kolejowe (miasta) w Europie, a krawędzie łączą

te wierzchołki, między którymi istnieje linia kolejowa.

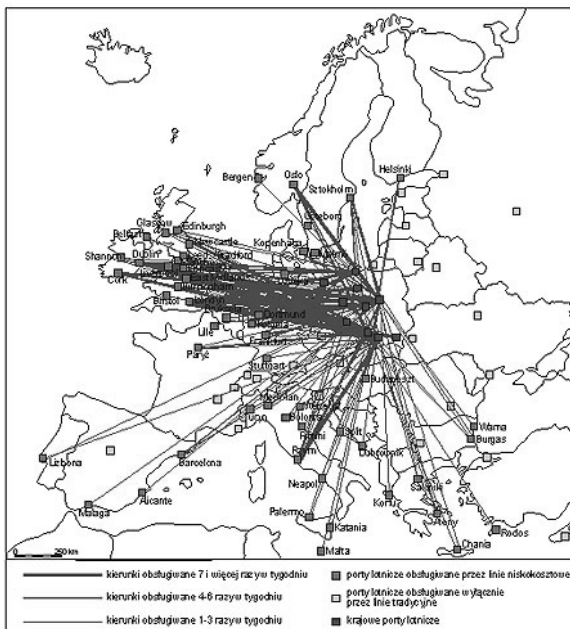


Rys. 9. Sieć szybkich kolei w Europie (stan na rok 2011); Źródło: [41]

sterować grubością krawędzi, możemy na krawędzi napisać liczbę jako wartość tej funkcji).



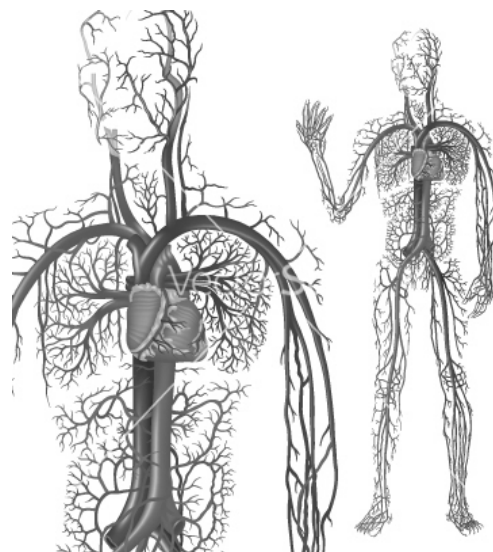
Rys. 11. Sieć połączeń lotniczych w USA; Źródło: [43]



Rys. 10. Sieć połączeń lotniczych liniami niskokosztowymi ze wszystkich polskich portów lotniczych (stan na czerwiec 2007); Źródło: [42]

3.3. Biologia

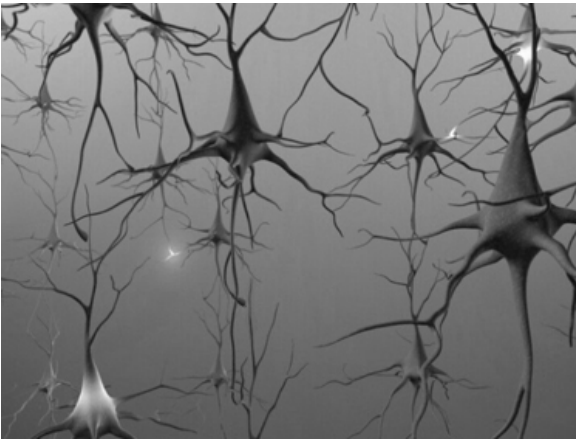
W tym rozdziale zawarto kilka przykładów dotyczących biologii. Na rysunku 12 pokazano sieć krwiobiegu człowieka. Wierzchołkami w tej sieci są przecięcia arterii krwiobiegu oraz ważne punkty w organizmie, a krawędziami odcinki żył i aort.



Rys. 12. Sieć krwiobiegu człowieka; Źródło: [44]

Na rysunkach 10 i 11 przedstawiono odpowiednio: sieć połączeń lotniczych liniami niskokosztowymi ze wszystkich polskich portów lotniczych (stan na czerwiec 2007) oraz sieć połączeń lotniczych w USA. Wierzchołkami w tej sieci są porty lotnicze (miasta), a krawędziami łączą te wierzchołki, między którymi odbywają się loty. Dodatkowo grubość krawędzi świadczy o częstotliwości lotów na danym kierunku (im grubsza krawędź, tym częstsze loty). Zwróćmy uwagę, że w związku z tym mamy do czynienia z siecią (jako specyficznym grafem), gdyż na krawędziach grafu opisano funkcję „częstość lotów” (zamiast

Na rysunku 13 pokazano sieć neuronów w mózgu. Wierzchołkami w tej sieci są neurony, a krawędziami tworzone są przez połączenia między neuronami w mózgu.



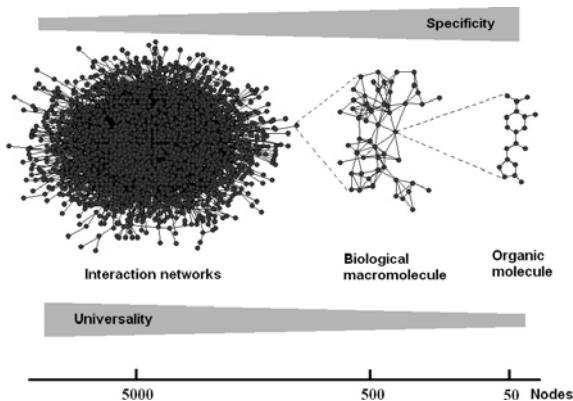
Rys. 13. Sieć neuronów w mózgu; Źródło: [45]

Na rysunku 14 przedstawiono sekwencję DNA człowieka. Wierzchołkami w tej sieci są proteiny, a krawędzie tworzą te pary protein, które są w interakcji ze sobą.



Rys. 14. Sekwencja DNA człowieka; Źródło: [46]

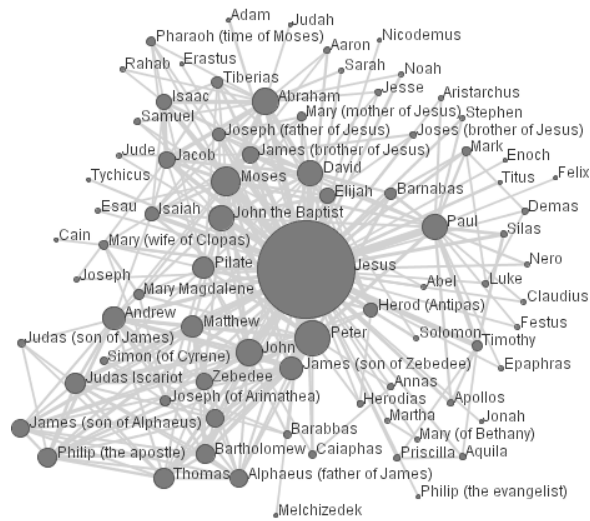
Na rysunku 15 przedstawiono biologiczne aspekty sieci złożonych. Jest to pewne uogólnienie rysunku 14, które pokazuje proces interakcji między molekułami od pojedynczej molekule aż do sieci oddziaływać wzajemnych.



Rys. 15. Biologiczne aspekty sieci złożonych; Źródło: [47]

3.4. Społeczeństwo

W tym rozdziale podano kilka przykładów dotyczących społeczeństwa i sieci społecznych. Na rysunku 16 pokazano sieć kontaktów (znajomości) między postaciami z Nowego Testamentu. Wierzchołkami są konkretne osoby, krawędzie łączą te pary wierzchołków (osób), które bezpośrednio się znały (pojawiała się informacja w Nowym Testamencie o tym, że kiedykolwiek te osoby się spotkały). Wielkość wierzchołka związana jest z liczbą znajomych tego wierzchołka (im większa liczba znajomych, tym większy wierzchołek).



Rys. 16. Sieć kontaktów (znajomości) między postaciami z Nowego Testamentu; Źródło: [48]

Rysunek 17 prezentuje sieć kontaktów (gier) między tenisistami. Wierzchołkami w tej sieci są tenisisti, którzy w latach 1968–2010 kiedykolwiek zajmowali pierwsze miejsce w rankingu ATP. Łuk w tym grafie łączy tę parę tenisistów (j, i), którzy kiedykolwiek ze sobą grali, a gracz i wygrał z graczem j (zwróćmy uwagę, że ważny jest zwrot łuku, który reprezentuje sytuację (przegrany \rightarrow wygrany)) [25]. Grubość łuku świadczy o liczbie przegranych gracza j grającego z graczem i (im grubszy łuk, tym większa liczba przegranych). Na tej podstawie autorzy pracy [25] wnioskujeją, kto był najlepszym tenisistą wszech czasów.

[14] opisanych poniżej, które mierzą pewne cechy charakterystyczne wierzchołków sieci, jednocześnie określając, który wierzchołek jest najważniejszy w sieci z punktu widzenia określonej miary.

Znormalizowany stopień dc_i (ang. *degree*) wierzchołka i :

$$dc_i = \frac{k_i}{N-1}, \quad (3)$$

gdzie: k_i oznacza stopień wierzchołka i w grafie G .

Najwyższą wartość tej miary uzyskuje wierzchołek, który ma największy stopień (a zatem ma najbardziej liczne sąsiedztwo).

Promień rc_i (ang. *radius*) wierzchołka i :

$$rc_i = \frac{1}{\max_{j \in V} d_{ij}}, \quad (4)$$

gdzie: d_{ij} – długość najkrótszej drogi w grafie G między wierzchołkami i oraz j (długość drogi w grafie między wierzchołkami i oraz j = liczba krawędzi na drodze z i do j).

Najwyższą ocenę uzyskuje wierzchołek, który jest możliwie najbliżej wszystkich najbardziej wysuniętych wierzchołków sieci (odległość dzieląca go od najdalszego wierzchołka jest najmniejsza).

Bliskość cc_i (ang. *closeness*) wierzchołka i :

$$cc_i = \frac{N-1}{\sum_{j \in V} d_{ij}}, \quad (5)$$

Wedle tej miary wierzchołek jest tym bardziej centralny, im jest średnio bliżej wszystkich innych wierzchołków sieci. W efekcie miara ta pozwala stwierdzić, który z dowolnych dwóch wierzchołków wymaga mniej kroków, aby „skomunikować” się z dowolnym innym wierzchołkiem sieci.

Pośrednictwo bc_i (ang. *betweenness, load*) wierzchołka i :

$$bc_i = \frac{\sum_{l \in V} \sum_{k \neq l \in V} \frac{p_{l,i,k}}{p_{l,k}}}{(N-2)(N-1)} \quad (6)$$

gdzie: $p_{l,i,k}$ – liczba dróg w grafie G między wierzchołkami l oraz k przechodzących przez i .

Miara ta definiowana jest jako ułamek liczby najkrótszych dróg łączących dwa dowolne wierzchołki sieci, które zawierają dany wierzchołek. Zazwyczaj wartości tej miary normalizuje się poprzez uwzględnienie maksymalnej możliwej liczby najkrótszych dróg w grafie będącym grafem pełnym⁷.

Współczynnik gronowania (klasteryzacji) gc_i (ang. *clusterization*) wierzchołka i :

$$gc_i = \frac{2E_i}{k_i(k_i-1)}, \quad k_i > 1, \quad (7)$$

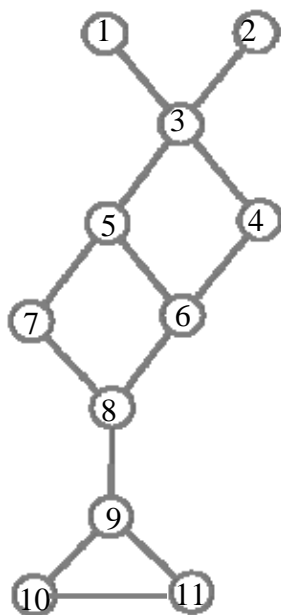
gdzie: E_i – liczba krawędzi między sąsiadami wierzchołka i .

Miara ta opisuje „prawdopodobieństwo”, że pierwsi (tzn. najbliżsi) sąsiedzi⁸ wierzchołka i są również swoimi pierwszymi sąsiadami. Okazuje się, że ta miara ma bardzo ciekawą interpretację socjologiczną. Podczas kształtowania się sieci społecznych dochodzi do efektu gronowania się (grupowania) ludzi o podobnych poglądach, zainteresowaniach itp. Na nasze stosunki z innymi osobami mogą wpływać również osoby trzecie. Z tego powodu socjologowie analizujący interakcje w sieciach społecznych zaczęli badać również stosunki między trzema osobami. Relacje wewnątrz takiego trójkąta są znacznie bardziej złożone i mogą prowadzić do ciekawych wniosków. Współczynnik gronowania odzwierciedla taką właśnie sytuację, w której zliczamy stosunek trójkątów⁹ występujących w sieci, do wszystkich trójkątów mogących potencjalnie występować. Przykładem takiego trójkąta jest podgraf zbudowany na wierzchołkach o numerach: 9, 10, 11 z rysunku 20.

⁷ Graf pełny to taki graf, w którym każdy wierzchołek połączony jest z każdym z pozostałych. W grafie o N wierzchołkach możemy mieć $N(N-1)$ takich połączeń.

⁸ Potocznie pierwszymi sąsiadami danego wierzchołka nazywa się te wierzchołki, które są z nim połączone bezpośrednio krawędzią (łukiem). Dalszymi (drugimi, trzecimi, itd.) sąsiadami nazywa się te wierzchołki, z którymi dany wierzchołek połączony jest pośrednio poprzez swoich sąsiadów (bliższych i innych dalszych).

⁹ Trójkąt w grafie (sieci) to taka jego część (podgraf) składająca się z trzech wierzchołków, w której każdy wierzchołek z każdym z pozostałych dwóch jest połączony krawędzią (łukiem). Inaczej mówiąc, jest to tzw. klika w grafie (podgraf, w którym każdy wierzchołek z kliki z każdym innym z kliki połączony jest krawędzią) o liczności 3.

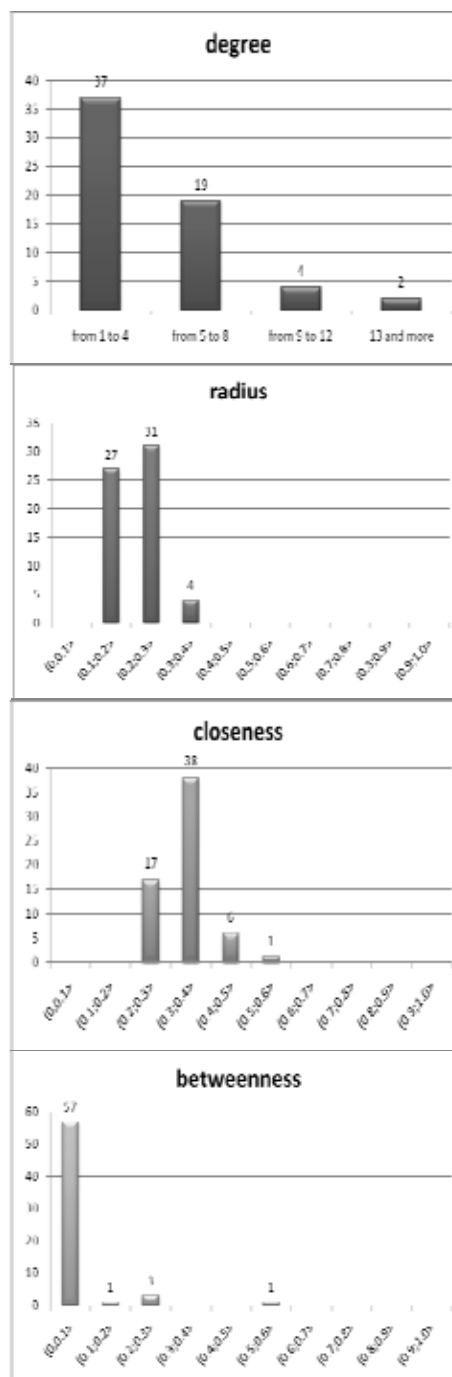


Rys. 20. Przykład grafu do badania miar centralności wierzchołków; Źródło: opracowanie własne

Niech graf z rysunku 20 opisuje pewną siatkę jedenastu znajomych (wierzchołkami są osoby, a krawędzie reprezentują bezpośrednie znajomości między osobami, np. wspólne miejsce pracy). Dla tej sieci otrzymamy następujące wierzchołki centralne (tzn. takie, dla których wartość miary jest największa) dla poszczególnych miar centralności:

- $dc_i - i = 3$; oznacza to, że osoba numer 3 ma najwięcej bezpośrednich znajomych z tej siatki
- $rc_i - i \in \{6, 7\}$; oznacza to, że osoby numer 6 i 7 potrzebują najmniej kontaktów (poprzez swoich znajomych, znajomych swoich znajomych itd.), żeby skomunikować się z najbardziej odległym znajomym w tej sieci
- $cc_i - i = 6$; oznacza to, że osoba numer 6 potrzebuje najmniej kontaktów w tej sieci, aby skomunikować się z każdą inną osobą w sieci
- $bc_i - i = 8$; oznacza to, że osoba numer 8 będzie najczęściej proszona o kontakt z jej znajomymi spośród innych osób tej sieci, tzn. aby dotrzeć do niektórych osób w sieci niezbędny będzie kontakt z osobą numer 8 (np. każda z osób 1, ..., 6 musi prosić osobę numer 8 o kontakt z osobą numer 9, dzięki której otrzymają kontakt do osób numer 10 i 11)
- $gc_i - i = 9$; oznacza to, że osoba numer 9 ma najwięcej znajomych, którzy również się znają (dodatkowo jest jedynym wierzchołkiem w sieci, którego znajomi (10 i 11) się znają).

Bardzo ciekawą analizę charakterystyk sieciowych dla sieci terrorystów z 11 września 2001 r. przedstawiono w pracy [29], rysunek 18. Wartości miar centralności (3)÷(6) wierzchołków tej sieci przedstawiono na rysunku 21. Widać wyraźnie, że w tej sieci istniały dokładnie dwa wierzchołki, których stopień był równy lub większy niż 13.



Rys. 21. Miary centralności wierzchołków dc , rc , cc , bc dla sieci terrorystów z 11 IX 2001 r.
Źródło: [29]

Charakterystyki poprzednio opisane dotyczyły wierzchołków grafu (sieci). Zdefiniujemy teraz dwie podstawowe charakterystyki dotyczące

sieci: średnią odległość i współczynnik gronowania (średni) sieci.

Średnia odległość L (średnia długości dróg najkrótszych) w sieci:

$$L = \frac{\sum_{i \neq j \in V} d_{ij}}{N(N-1)}. \quad (8)$$

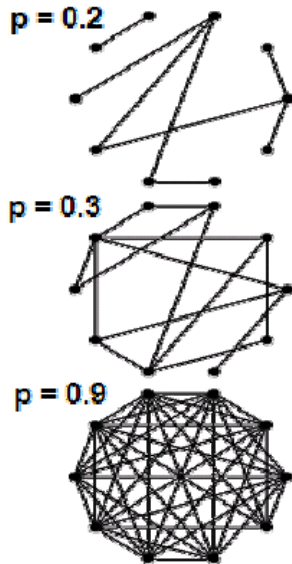
Współczynnik gronowania (średni) C sieci:

$$C = \frac{1}{N} \sum_{i \in V} g c_i. \quad (9)$$

Wykorzystanie tych charakterystyk pokazano w dalszych rozdziałach.

4.2. Grafy losowe

W roku 1959 P. Erdős i A. Renyi zaproponowali model grafu losowego (model ER) [10]. Model ten charakteryzuje się tym, że każda para wierzchołków (spośród N) łączona jest krawędzią z prawdopodobieństwem p . Przykłady grafów o $N = 10$ wierzchołkach, dla różnych wartości prawdopodobieństwa p , przedstawia rysunek 22.



Rys. 22. Przykłady grafów losowych o $N = 10$ wierzchołkach dla różnych wartości p ; Źródło: [10]

Rozkład $P(k)$ stopnia k -tego wierzchołka w takim modelu jest najczęściej dwumianowy:

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k} \quad (10)$$

lub Poissona:

$$P(k) \cong \frac{(pN)^k}{k!} e^{-pN} = \frac{\bar{k}^k}{k!} e^{-\bar{k}}, \quad (11)$$

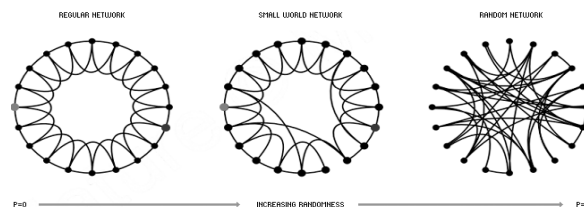
gdzie: \bar{k} – średnia stopni wierzchołków w grafie. Podsumowując, graf losowy ma następujące cechy:

- niski współczynnik gronowania C
- małą średnią wartość długości dróg L .

4.3. Sieci małego świata

Wiele pokoleń naukowców dorastało w przekonaniu, że deterministyczne grafy regularne¹⁰, a później grafy losowe są najbardziej naturalnymi modelami układów złożonych o strukturze sieci. Okazało się, że wyniki badań wskazują, iż sieci rzeczywiste mają zupełnie inne własności niż grafy regularne, czy też grafy losowe. Są to tzw. sieci złożone [14].

W roku 1998 D. Watts i S. Strogatz opublikowali pracę [33], w której zaproponowali model „małego świata” (ang. *Small World*, W-S, mały świat). Model W-S powstaje z sieci regularnej, w której losowo wybrane krawędzie są „przepinane” z prawdopodobieństwem p . W związku z tym w takiej sieci pojawiają się krótsze drogi między odległymi parami wierzchołków (rysunek 23).



Rys. 23. Przykłady sieci (w kolejności od lewej): sieć regularna, sieć typu *Small World*, sieć losowa; Źródło: [33]

Średnia odległość (średnia długość dróg definiowana przez (8)) między dwoma dowolnymi wierzchołkami jest w tych sieciach mała w porównaniu z liczbą wierzchołków (np. dla sieci Internet¹¹ wynosi 3.31; dla sieci

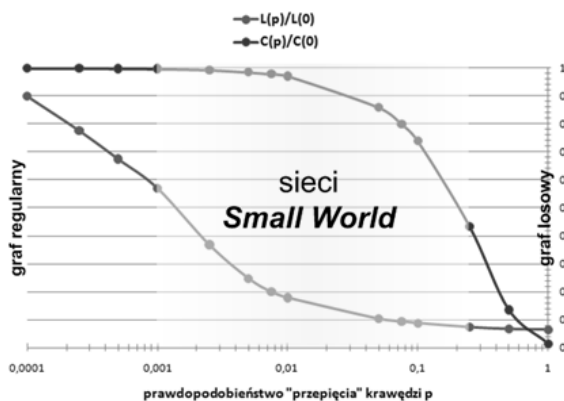
¹⁰ Grafy regularne to takie, w których stopień każdego wierzchołka ma tę samą wartość (ewentualnie bardzo zbliżoną).

¹¹ Wierzchołkami w tej sieci są strony www, a krawędzią łączymy stronę www z tymi stronami, do których są linki na rozpatrywanej stronie.

współpracy naukowej¹² 6.2) i słabo zależy od rozmiaru sieci:

$$L \sim \frac{\ln N}{\ln k} \quad (12)$$

Zgodnie ze spostrzeżeniami Watts'a i Strogatza, sieć może być uważana za sieć małego świata wtedy, gdy średnia odległość L w tej sieci jest porównywalna ze średnią odległością w sieci losowej oraz posiada znacznie większą wartość średniego współczynnika gronowania C .



Rys. 24. Sieci *Small World* na tle sieci regularnych i losowych; Źródło: [19]

Na rysunku 24 pokazano zależność między sieciami typu *Small World* a sieciami regularnymi i losowymi. Funkcje $L(p)$ i $C(p)$ oznaczają odpowiednio średnią długość dróg w sieci oraz współczynnik klasteryzacji (gronowania) sieci przy prawdopodobieństwie p „przejęcia” krawędzi ($L(0)$, $C(0)$ oznaczają wartości tych charakterystyk dla grafu regularnego). Widać wyraźnie, że dla sieci typu *Small World*, przy $p \in (0.001; 0.1)$, mamy niską średnią długość dróg, przy współczynniku gronowania pozostającym na wciąż wysokim poziomie (w stosunku do grafu regularnego).

Podsumowując, sieci typu *Small World* mają następujące cechy:

- wysoki współczynnik gronowania
- małą średnią wartość długości dróg.

Z opisanymi własnościami sieci typu *Small World* wiąże się tzw. reguła sześciu stopni separacji (ang. *six degrees of separation*). Bazuje ona na spostrzeżeniu, że w rzeczywistych sieciach społecznych średnia długość drogi L wynosi około 6. Reguła ta przejawia się

w potocznym stwierdzeniu, że świat jest mały: do uściśnięcia dłoni dowolnego człowieka na świecie potrzeba średnio sześciu uścisków. Potwierdzona została eksperymentem socjologa Stanleya Milgrama z 1967 r. [32]. Wysłał on zwykłą pocztą do 160 przypadkowo wybranych osób list zawierający wyjaśnienie eksperymentu, zdjęcie i adres pewnej osoby; jeśli adresat znał osobiście człowieka wymienionego w liście, miał przesłać list bezpośrednio do niego; jeżeli nie znał, to list miał być przesłany do innego znajomego, o którym adresat sądził, że może znać poszukiwaną osobę lub przynajmniej znać kogoś, kto ją zna osobiście. W 2006 r. E. Horvitz i J. Leskovec z Microsoftu przeanalizowali wirtualne trasy 30 mld wiadomości, które zostały wysłane z 240 mln komputerów z całego świata w czerwcu 2006 r. [21]. W obu eksperymentach okazało się, że średnia długości dróg przesyłek wyniosła ok. 6.

W dniu 23 listopada 2011 r. brytyjski *DailyMail* opublikował artykuł [49], w którym podał przykład z rysunku 25.



Rys. 25. Od gwiazdy pop do dyktatora – praktyczny przykład reguły sześciu/pięciu stopni separacji; Źródło: [49]

Tytuł w przykładzie jest dość znamienny: od gwiazdy pop do dyktatora. Uzasadniano w nim opisaną wyżej własność sieci *Small World*. Gwiazda muzyki pop, Charlotte Church, w programie *Somewhere over the Rainbow* poszukiwała aktorki (wykonawczynie) do spektaklu muzycznego Andrew Lloyda Webera (1), którą to rolę zagrała Scarlett Johansson (2), ta z kolei grała w filmie *Lost in Translation* z Billem Murrayem (3), który zagrał w komedii *Space Jam* razem z gwiazdą NBA Michaeliem Jordanem (4), który podarował własnoręcznie podpisaną piłkę do koszykówki niezwiązanej już dyktatorowi Korei Północnej Kim Dzong Ilowi (5). Można zatem powiedzieć, że gwiazdę muzyki pop Charlotte Church dzieli od dyktatora Kim Dzong Ila tylko pięć kontaktów (pięć uścisków dłoni).

Jeżeli chodzi o świat aktorski, to takim „modelowym” aktorem, na przykładzie którego

¹² Wierzchołkami w tej sieci są naukowcy, a krawędzie łączą tych naukowców, którzy mają wspólne publikacje.

pokazuje się zasadę sześciu stopni separacji, jest Kevin Bacon [14].

4.4. Sieci bezskalowe

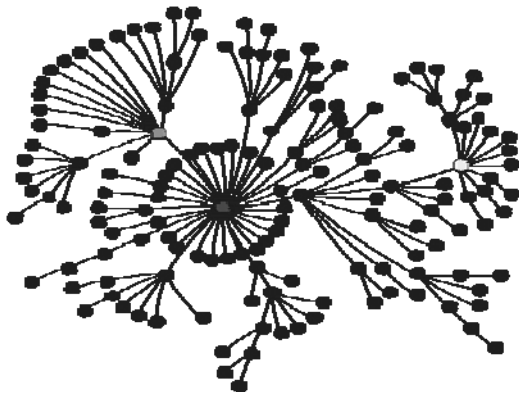
W roku 1999 A. Barabasi i R. Albert opublikowali przełomową pracę [3], w której zaproponowali model sieci „bezskałowej” (ang. *Scale Free*, B–A) przedstawiony na rysunku 26. Model B–A sieci *Scale Free* charakteryzuje się:

- stałym wzrostem rozmiaru sieci (ewolucja, spontaniczność)
- preferencyjnym dołączaniem nowych wierzchołków do sieci
- potęgowym rozkładem stopnia (k) wierzchołka (rysunek 28):

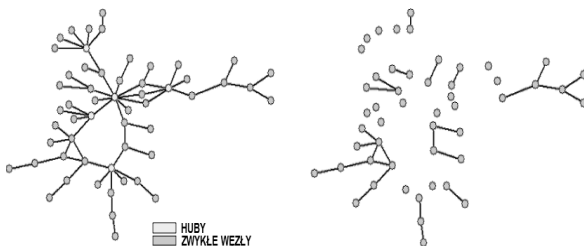
$$P(k) \sim k^{-\gamma}, \quad (13)$$

gdzie: γ jest parametrem rozkładu, np. dla sieci internetowej $\gamma \cong 2.5$, dla sieci cytowań artykułów naukowych z tzw. listy filadelfijskiej $\gamma \cong 3.15$, dla sieci mocy trzęsień ziemi $\gamma \cong 1.88$ [14]. W praktyce parametr γ w sieciach rzeczywistych w zasadzie nie przekracza wartości 3.

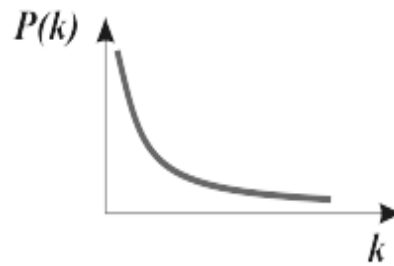
Potęgowy rozkład stopnia wierzchołka jest podstawową cechą tych sieci i decyduje o tym, że **większość wierzchołków ma niski stopień, ale niektóre mają bardzo duży stopień** (tzw. huby, tzn. wierzchołki, które mają największy stopień na rysunku 27).



Rys. 26. Sieć *Scale Free*; Źródło: [19]



Rys. 27. Huby w sieci *Scale Free* (po lewej) oraz sieć po usunięciu hubów (po prawej); Źródło: [19]



Rys. 28. Funkcja gęstości potęgowego rozkładu prawdopodobieństwa; Źródło: opracowanie własne

Dlaczego o sieciach tych mówi się, że są bezskalowe? Bezskałowość to inaczej niemierzalność: obiekty w każdej skali wyglądają tak samo, a decyduje o tym prawo potęgowe opisujące ich strukturę (np. fraktale: linia brzegowa Wielkiej Brytanii jest fraktalem i jej długości nie można dokładnie zmierzyć, ponieważ dobór odpowiedniej skali pomiaru jest niemożliwy). Dodatkowo przeanalizujemy dokładnie zależność (14) opisującą rozkład potęgowy i jego podstawowe charakterystyki: wartość oczekiwaną i wariancję. Z tej analizy wynika, że wariancja ciągłej zmiennej losowej o rozkładzie potęgowym dla parametru $\gamma < 3$ jest nieskończona; podobnie zachowuje się wartość oczekiwana, z tym że dla $\gamma < 2!$

$$X \sim P(x) = cx^{-\gamma}$$

$$EX = \int_{x_{\min}}^{\infty} xP(x)dx = c \int_{x_{\min}}^{\infty} x^{-\gamma+1} dx \quad (14)$$

$$V^2 X = \int_{x_{\min}}^{\infty} x^2 P(x)dx = c \int_{x_{\min}}^{\infty} x^{-\gamma+2} dx$$

gdzie:

X – zmienna losowa o rozkładzie potęgowym $P(x)$;

EX – wartość oczekiwana zmiennej losowej X ;

$V^2 X$ – wariancja zmiennej losowej X .

Ponieważ, jak wcześniej napisano, w praktyce sieci rzeczywistych parametr γ nie przekracza wartości 3, mamy odpowiedź na pytanie dotyczące bezskałowości tych sieci.

Preferencyjne dołączanie (i oczywiście potęgowy rozkład stopnia wierzchołka) decyduje o tym, że nowe wierzchołki dołączają się z większym prawdopodobieństwem do tych, które mają większy stopień („bogaci stają się coraz bogatsi”).

Jedną z podstawowych cech sieci *Scale Free*, i co za tym idzie wszystkich sieci rzeczywistych wykazujących tę własność, jest ich stała ewolucja. Problemy ewolucji sieci

złożonych poruszono między innymi w pracach [4], [5], [19] oraz rozdziale 5.

Własności sieci *Scale Free* wyraźnie widać na wszystkich rysunkach 4÷18 z rozdziału 3.

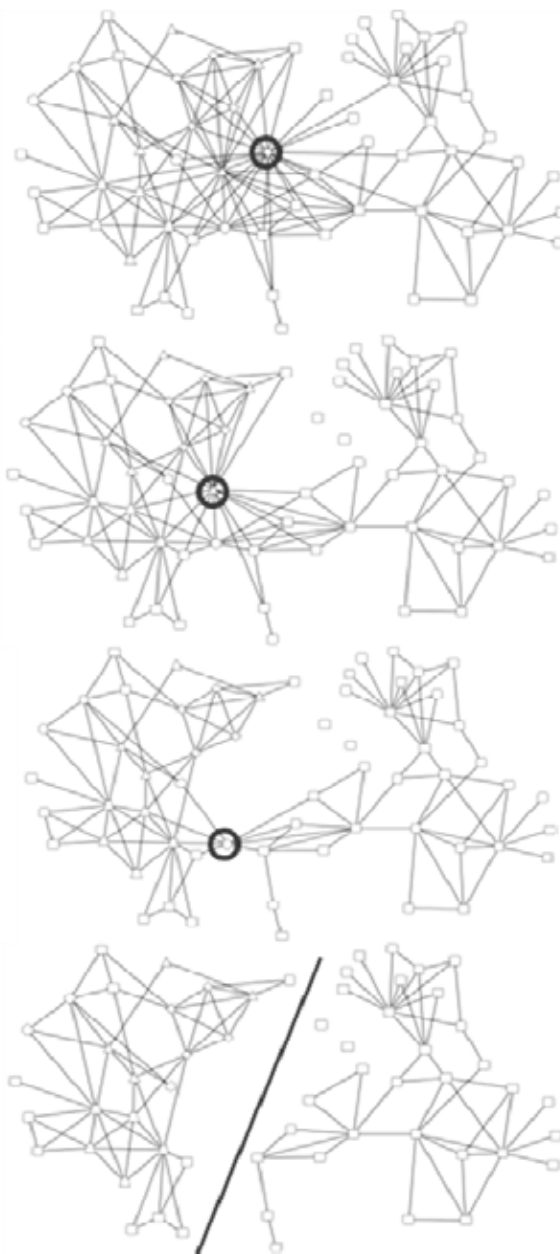
4.5. Odporność i podatność na ataki oraz propagację informacji w sieciach złożonych

Bezskalowość jest jedną z wielkich zalet sieci *Scale Free*, ale jednocześnie jedną z jej podstawowych wad.

Sieć typu *Scale Free* jest bardzo odporna na ataki losowe: np. dla sieci internetowej należałoby usunąć aż ok. 99% węzłów, aby rozspójnić¹³ taką sieć. Z kolei jest bardzo nieodporna na ataki celowane: np. dla sieci internetowej należałoby usunąć tylko ok. 6% najlepiej usieciowionych¹⁴ wierzchołków, aby ją rozspójnić.

Jakie można z tego wysnuć wnioski? Otóż, jeżeli sieć rzeczywista opisuje sieć społeczną, w której rozprzestrzenia się epidemia, to najskuteczniejszym sposobem zapobiegania dalszemu rozprzestrzenianiu się epidemii jest szczepienie tzw. hubów, czyli najlepiej usieciowionych wierzchołków, bo to poprzez nie odbywa się większość kontaktów w sieci. Zatem szczepiąc huby, najlepiej chronimy populację [17], [18], [19].

Jeżeli z kolei rozpatrujemy tę sieć z punktu widzenia propagacji informacji (plotki, opinii), to jest ona szybka, ale jednocześnie szybka jest propagacja błędu, wirusa, awarii [29].



Rys. 29. Eliminacje trzech najlepiej usieciowionych wierzchołków, aby rozspójnić sieć z rysunku 18; Źródło: [29]

Bardzo ciekawa analiza została przedstawiona na rysunku 29, a dotyczy sieci z rysunku 18 (sieć terrorystów, którzy dokonali ataku 11 IX 2001 r.). Z analizy tej wynika, że jeżeli eliminowalibyśmy z tej sieci po kolei wierzchołki o najwyższym stopniu (rysunek 21), to po trzech eliminacjach spowodowalibyśmy rozspójnienie tej sieci. Sieć podzielona zostałaby na dwie podsieci niemogące się komunikować ze sobą. Jest to bardzo ważne spostrzeżenie i wskazówka wykorzystywana między innymi w systemach wczesnego ostrzegania przed zagrożeniami np. terrorystycznymi [1].

¹³ Rozspójnienie sieci polega na tym, że po rozspójnieniu w takiej sieci będzie istniała przynajmniej jedna para wierzchołków, między którymi nie istnieje droga.

¹⁴ Najlepiej usieciowione węzły to te, które mają największe wartości miar centralności (kilku bądź jednej wybranej).

4.6. Algorytmiczne aspekty sieci złożonych

Pisząc o problemach sieci, nie sposób nie odnieść się do problemów algorytmicznych, w tym złożoności obliczeniowej. Z reguły sieci rzeczywiste są sieciami wielkich rozmiarów (miliony wierzchołków, np. sieć stron www), gdzie problemy złożoności i efektywnych algorytmów są szczególnie istotne. Przedstawimy oszacowanie złożoności obliczeniowej algorytmów wyznaczających wartości podstawowych miar sieciowych zaprezentowanych w rozdziale 4.1, posługując się przy tym symbolem notacji asymptotycznej O .

Do wyznaczenia wartości miary *degree* (3) dla wierzchołka i -tego musi wyznaczyć k_i w czasie $O(N)$, stąd wyliczenie wartości miary dla pojedynczego wierzchołka zajmuje czas proporcjonalny do $O(N)$. Dla wszystkich N wierzchołków zajmuje to czas proporcjonalny do $O(N^2)$.

Do wyznaczenia wartości miary *radius* (4) dla i -tego wierzchołka musimy wyznaczyć drzewo najkrótszych dróg o korzeniu w wierzchołku i , aby wyliczyć d_{ij} dla każdego $j \in V$. Zajmuje to czas proporcjonalny do $O(N+M)$, gdzie M to liczba krawędzi w grafie, wykorzystując algorytm BFS¹⁵. Następnie musimy wyznaczyć największą z wartości d_{ij} – zajmuje to czas proporcjonalny do $O(N)$. Stąd wyznaczenie wartości tej miary dla wszystkich N wierzchołków zajmuje czas proporcjonalny do $O(2N^2 + NM)$.

Do wyznaczenia wartości miary *closeness* (5) dla i -tego wierzchołka wykorzystujemy czas $O(N^2 + NM)$ (z obliczeniowego punktu widzenia operacja „max” zastępowana jest przez operację „sum”, reszta tak samo jak w procedurze wyznaczania wartości miary *radius*).

Szybki algorytm do wyznaczania wartości miary *betweenness* (6) został zaproponowany przez Brandesa [7] (szybszy niż w [13]) – dla pojedynczego wierzchołka potrzebuje czasu proporcjonalnego do $O(M)$. Stąd wyliczenie wartości tej miary dla wszystkich N wierzchołków zajmuje czas $O(NM)$.

Aby wyliczyć wartość współczynnika gronowania (*clusterization*) (7) dla i -tego wierzchołka, musimy w najgorszym przypadku rozpatrzeć $N-1$ sąsiadów tego wierzchołka, z których każdy może mieć $N-1$ sąsiadów. Zatem zajmuje to czas proporcjonalny do $O(N^2)$.

Ponieważ wyliczamy wartość tej miary dla każdego z N wierzchołków sieci, to otrzymujemy $O(N^3)$.

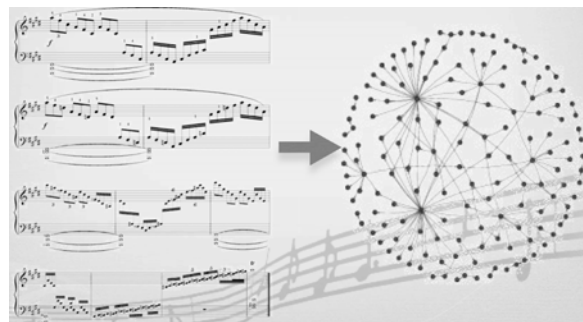
Złożoność wyliczenia charakterystyk sieciowych opisanych formułami (8) i (9) zależy od tego, czy wcześniej wyznaczyliśmy charakterystyki *radius* bądź *closeness* i *clusterization*.

Jeżeli wyliczyliśmy wcześniej wartości miary *closeness* bądź *radius* dla wszystkich wierzchołków, to wyliczenie charakterystyki opisanej formułą (8) zajmie czas proporcjonalny do $O(N)$, jeżeli nie wyliczyliśmy jej wcześniej – to $O(N^2 + NM)$.

Jeżeli wyliczyliśmy wcześniej wartości miary *clusterization* dla wszystkich wierzchołków, to wyliczenie charakterystyki opisanej formułą (9) zajmie czas proporcjonalny do $O(N)$, jeżeli nie wyliczyliśmy jej wcześniej – to $O(N^3 + N)$.

5. Dalsze przykłady zastosowań sieci złożonych

W tym rozdziale przedstawimy uzupełnienie rozdziału 3 dotyczące przykładów zastosowań sieci złożonych. Zdaniem autora przykłady te mogą zaskoczyć Czytelnika swoją oryginalnością i jednocześnie dziedziną, której dotyczy, dowodząc jednocześnie w zasadzie nieograniczonych zastosowań modeli grafowo-sieciowych.



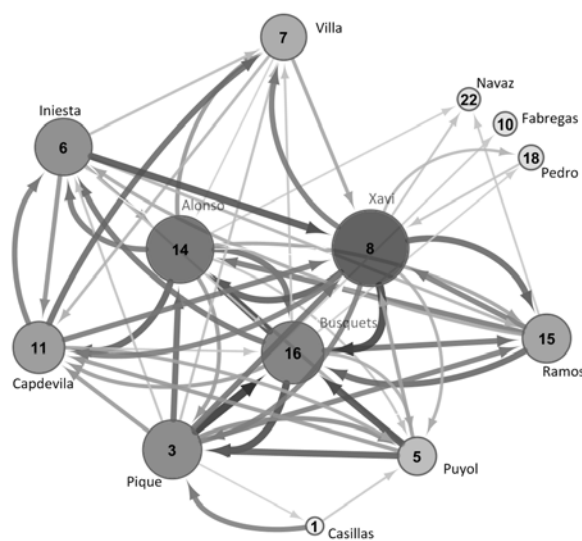
Rys. 30. Tworzenie muzyki z wykorzystaniem sieci złożonych;
Źródło: [22]

Na rysunku 30 przedstawiono przykład modelowania i analizy utworu muzycznego (jeden z utworów J.S. Bacha) z wykorzystaniem sieci złożonej [22]. Wierzchołkami tego grafu są poszczególne nuty, a krawędzie łączą te nuty, które współlistnieją obok siebie w linii melodycznej utworu. Nowe aranżacje znanych utworów muzycznych, często generowane komputerowo, powstają w ten sposób, że mając

¹⁵ Best First Search.

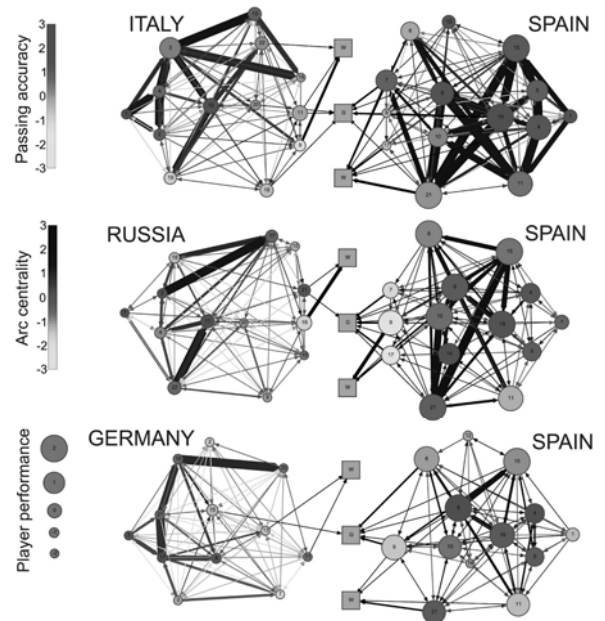
sieć złożoną reprezentującą utwór, dokonujemy dowolnych przekształceń w tej sieci (np. „przepinając” jakieś krawędzie z odpowiednim prawdopodobieństwem tak, jak przy tworzeniu sieci *Small World* z sieci regularnej).

Bardzo ciekawy przykład zastosowania sieci złożonych dotyczy analizy sportowych gier zespołowych. Na rysunku 31 przedstawiono sieć kontaktów (podań) reprezentacji Hiszpanii w piłce nożnej na mundialu w RPA 2010 [9], [23]. Wierzchołkami w tej sieci są piłkarze, łuk łączy parę piłkarzy (i, j) wówczas, gdy piłkarz i podawał piłkę do piłkarza j . Wielkość wierzchołka w sieci zależy od jego istotności, tzn. od liczby podań, która trafia do/od danego piłkarza. Grubość łuku zależy od liczby podań wymienianych między parą piłkarzy: im grubszy łuk, tym więcej podań między daną parą piłkarzy (tym łuk istotniejszy). Z rysunku 31 wyraźnie widać, który z piłkarzy jest najważniejszy z punktu widzenia prowadzenia gry (8 – Xavi; 14 – Alonso; 16 – Busquets) oraz między którymi piłkarzami najczęściej ta gra się odbywa (Iniesta → Xavi, Xavi → Busquets, Pique → Busquets). Znamienne jest to, że od pięciu lat ta taktyka się praktycznie nie zmieniła, ale żadnej z drużyn nie udało się zatrzymać reprezentacji Hiszpanii, która wygrała wszystkie najważniejsze turnieje w tym okresie (dwa razy mistrzostwa europy, jeden raz mistrzostwa świata). Jednym z powodów takiego stanu rzeczy jest fakt, że w drużynie Hiszpanii istnieją co najmniej 4 wierzchołki centralne (a nie jeden, jak w większości drużyn), które trudno jednocześnie „izolować” przeciwnikowi w tej sieci.



Rys. 31. Sieć kontaktów (podań) reprezentacji Hiszpanii w piłce nożnej na mundialu w RPA 2010; Źródło: [9]

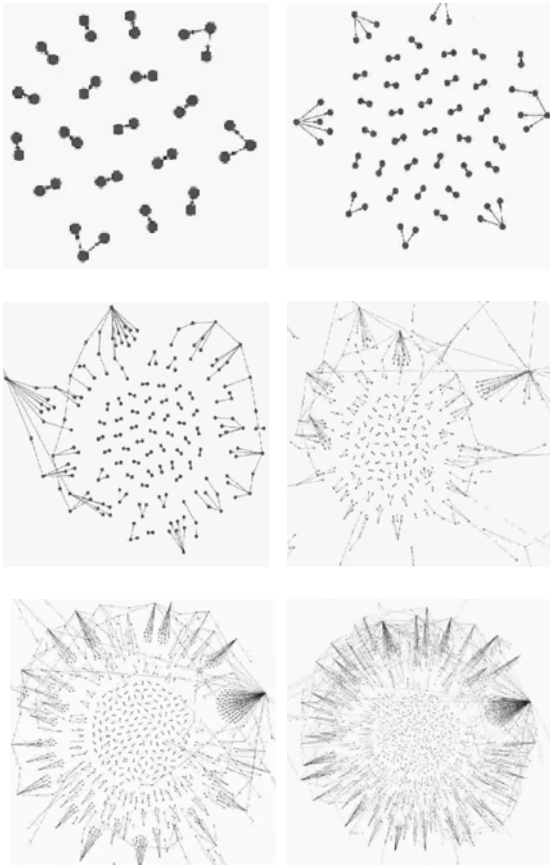
Na rysunku 32 przeanalizowano, według podobnego schematu jak na rysunku 31, reprezentację Hiszpanii w czasie meczów na mundialu w RPA 2010 z reprezentacjami: Włoch, Rosji i Niemiec. Widać wyraźnie różnice w stylu gry tych czterech reprezentacji. Z analizy tych sieci wynika, że najbardziej podobny styl (z dużą wymianą podań, kilkoma wierzchołkami centralnymi) do stylu reprezentacji Hiszpanii prezentuje reprezentacja Włoch.



Rys. 32. Sieć kontaktów (podań) reprezentacji Hiszpanii oraz trzech innych reprezentacji grających z Hiszpaniami na mundialu w RPA 2010; Źródło: [9]

W świecie futbolowym znanym wielkim zwolennikiem wykorzystania metod ilościowych (w tym również sieciowych) do analiz boiskowych, „rozpracowywania” przeciwnika i ustalania najlepszej taktyki boiskowej jest Rafael Benitez, były trener drużyny futbolowej FC Liverpool (za czasów reprezentowania barw tego klubu przez polskiego bramkarza Jerzego Dudka).

Jedną z cech charakterystycznych sieci współczesnych jest ich dynamiczny charakter i stały rozwój. Na rysunku 33 pokazano przykład ewolucji sieci rzeczywistej (z około jednej godziny), dotyczący ewolucji komunikatów na Twitterze związanych z komentowaniem rezygnacji z urzędu prezydenta Egiptu H. Mubaraką, w dniu 11 lutego 2011 r. Wierzchołkami tej sieci są komunikaty (czyli konkretne osoby, które je umieściły), a krawędzie łączą te osoby, z których co najmniej jedna skomentowała wpis tej drugiej.



Rys. 33. Przykład ewolucji komunikatów na Twitterze dotyczących komentowania rezygnacji prezydenta Egiptu H. Mubarak z urzędu w dniu 11 lutego 2011 r.; Źródło: [52]

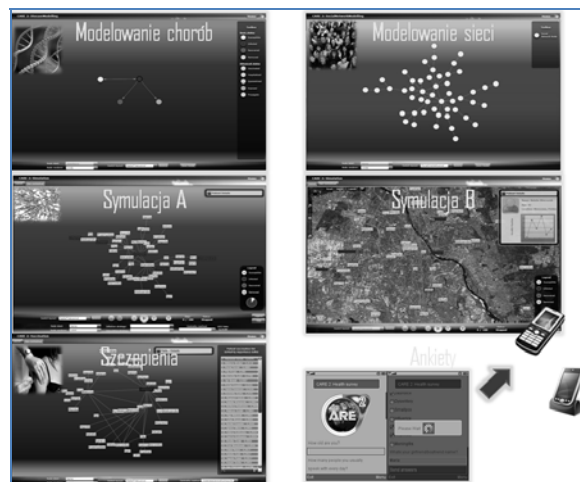
6. Przykłady modeli sieciowych wykorzystywanych w projektach realizowanych na Wydziale Cybernetyki WAT

W poprzednim rozdziale pozwoliliśmy sobie na pokazanie kilku zastosowań modeli i metod grafowo-sieciowych, które z nauką miały mało wspólnego. W tym rozdziale zostaną zaprezentowane wybrane przykłady projektów naukowo-badawczych i wdrożeniowych realizowanych na Wydziale Cybernetyki WAT, w których wykorzystano modele i metody teorii grafów i sieci.

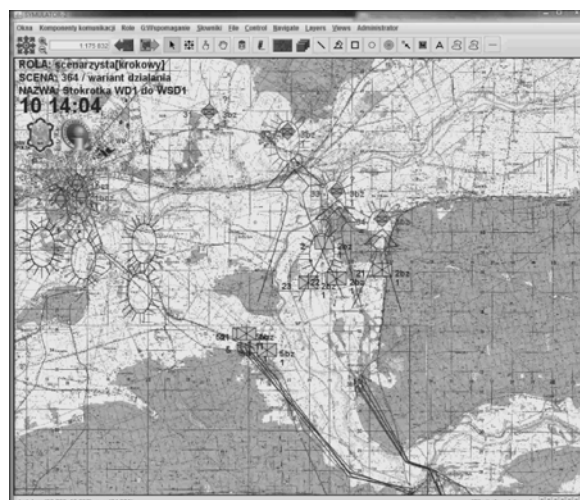
CARE2 (ang. *Creative Application to Remedy Epidemics*), przedstawiony na rysunku 34, to symulacyjny system wspomagania podejmowania decyzji do walki z rozprzestrzenianiem się chorób zakaźnych w sieciach społecznych [18]. Funkcjonalność systemu została podzielona na następujące moduły: Modelowanie Chorób Zakaźnych, Modelowanie Sieci Społecznych, Zbieranie Danych o Topologii Sieci Społecznych,

Symulacja Epidemii oraz Konstrukcja Strategii Szczepień.

System jest narzędziem wspierającym działanie służb epidemiologicznych i/lub centrów zarządzania kryzysowego, daje możliwości prognozowania dynamiki epidemii, szacowania środków medycznych potrzebnych do walki z epidemią oraz optymalizacji strategii szczepień. Wiedzę o strukturze powiązań, jaką posiadamy w przypadku większości sieci społecznych, możemy określić jako wysoce niepełną i niepewną. Cechą szczególną systemu jest zdolność do budowy efektywnej strategii szczepień również przy braku precyzyjnych danych o populacji, dzięki wykorzystaniu specjalizowanych kwestionariuszy osobowych.

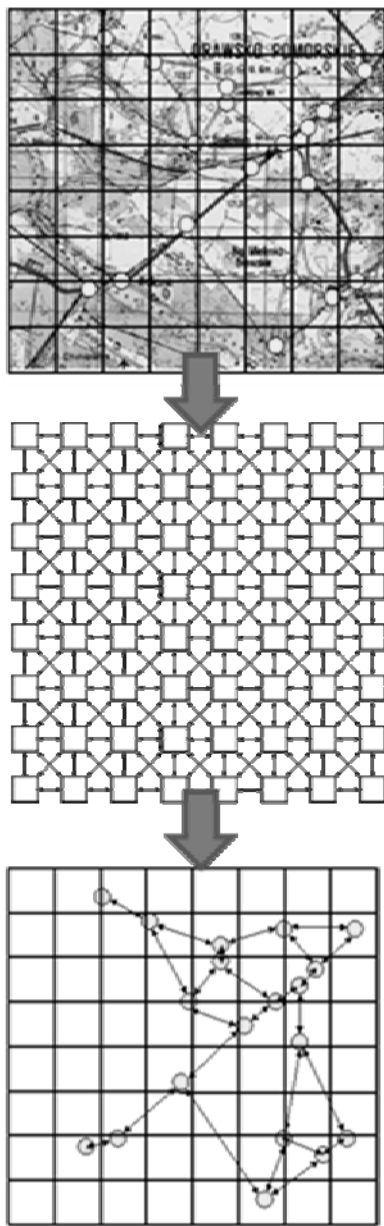


Rys. 34. Struktura modułowa systemu CARE2 do modelowania i symulowania rozprzestrzeniania się epidemii w sieciach społecznych; Źródło: [18]



Rys. 35. Okno główne symulatora deterministycznego CAVARS; Źródło: [2], [51]

Na rysunku 35 przedstawiono okno główne symulatora deterministycznego CAVaRS (ang. *Course of Action Verification and Recommendation Simulator*) [2], który powstał na Wydziale Cybernetyki WAT i został wdrożony na Wydziale Zarządzania i Dowodzenia Akademii Obrony Narodowej dla potrzeb ćwiczeń sztabowych. W symulatorze tym, podobnie jak w symulatorze stochastycznym *Złocień* z rysunku 37 ([30], [51]), zastosowano modele sieciowe do reprezentacji środowiska pola walki (rysunek 36) oraz algorytmy grafowo-sieciowe na potrzeby m.in. planowania przemieszczania.

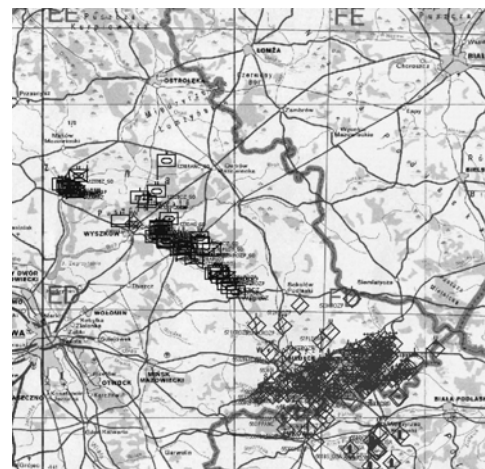


Rys. 36. Grafowo-sieciowe modele terenu w postaci sieci kwadratów terenu i sieci drogowo-kolejowej w systemach *Złocień* i CAVaRS;
Źródło: [30]

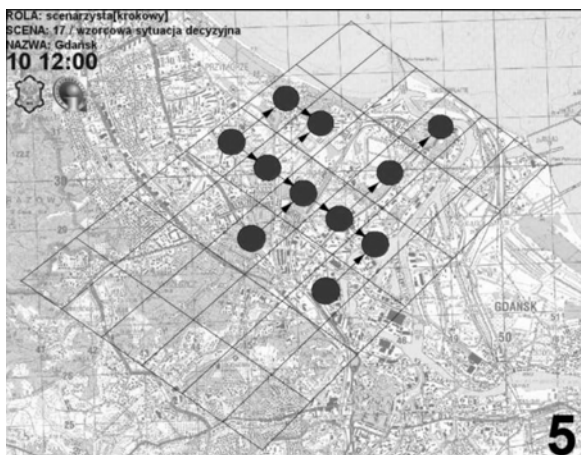
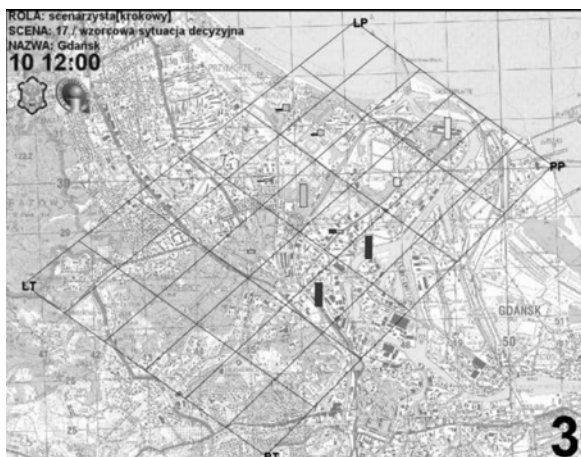
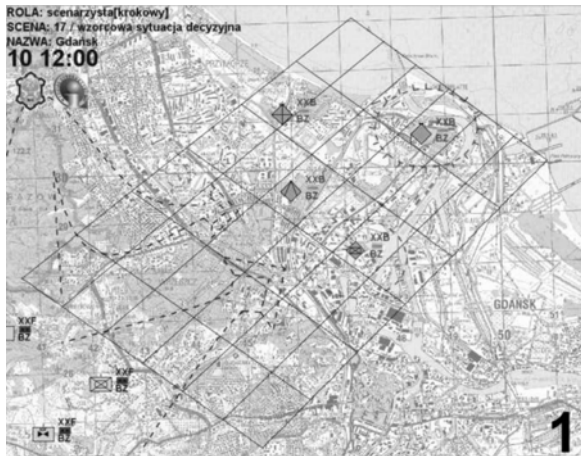
Model środowiska pola walki (terenu) oparty jest o mapę cyfrową w formacie VPF i jest dwoisty: (1) w postaci regularnej sieci kwadratów terenu, (2) w postaci sieci drogowo-kolejowej. Model w postaci regularnej sieci kwadratów terenu jest używany do planowania przemieszczania jednostek „na przełaj”, np. w czasie natarcia. Drugi z modeli używany jest do planowania szybkich przemieszczeń, np. marszów po rzeczywistych drogach. Oba te modele są zintegrowane. Integracja ta polega na tym, że istnieje możliwość przejścia z jednego modelu do drugiego i na odwrót. Charakterystyki obu modeli terenu zależą od: topografii terenu, pogody, pory doby oraz roku, oddziaływania wojsk własnych oraz przeciwnika (np. zniszczenie mostu będącego elementem sieci drogowo-kolejowej).

Bazując na modelach terenu modele przemieszczania definiują następujące plany przemieszczania: z punktu (rejonu) do punktu (rejonu); przechodzące przez wybrane punkty (rejonu); omijające wybrane punkty (rejonu, przeszkody); wewnątrz lub na zewnątrz pewnego rejonu; po drogach i (lub) „na przełaj”. Plany przemieszczania mogą być oparte o następujące kryteria: minimalizacja czasu przemieszczania, minimalizacja długości drogi, maksymalizacja stopnia maskowania.

Grafowo-sieciowe algorytmy planowania przemieszczania realizują wspomniane modele przemieszczania i wykorzystują zmodyfikowane algorytmy wyznaczania dróg ekstremalnych: algorytm Dijkstry, algorytm Johnsona dla rzadkich sieci, algorytm A*, algorytmy wyznaczania najkrótszych dróg geometrycznych.



Rys. 37. Okno główne symulatora stochastycznego *Złocień* w czasie symulacji marszu brygady;
Źródło: [30], [51]



Rys. 38. Przykład działania grafowo-sieciowego modułu identyfikacji sytuacji decyzyjnych w systemie CAVaRS; Źródło: [12]

Na rysunku 38 zaprezentowano przykład działania grafowo-sieciowego modułu identyfikacji sytuacji decyzyjnych w systemie CAVaRS. Moduł ten służy do automatycznego rozpoznawania sytuacji decyzyjnej na polu walki i rekomendowania dowódcy najlepszego (z punktu widzenia wielu kryteriów) wariantu działania dla aktualnej sytuacji. Oparty jest o metodę wielokryterialnego podobieństwa grafów ważonych (sytuacje decyzyjne

reprezentowane są za pomocą grafów ważonych) [26], [27].

Szczegółowy opis wybranych projektów realizowanych na Wydziale Cybernetyki WAT, w których wykorzystywano metody grafowo-sieciowe, dostępny jest pod adresem [51].

7. Wnioski

Sieci są wszechobecne w świecie. Wiele ich zastosowań zostało przytoczonych, ale oczywiście nie wszystkie. Oto niektóre z wymienianych zastosowań:

- Biologia:
 - Sieć połączeń neuronowych w mózgu
 - Interakcje między proteinami
 - Sieć transportu wody w roślinach
 - Krwiobieg
 - Struktury DNA, ...
- Socjologia, społeczeństwo:
 - Sieci kontaktów międzyludzkich (serwisy społecznościowe (Facebook, Twitter), organizacje, partie, zrzeszenia, współpraca naukowa)
 - Sieci przestępcze
 - Sojusze w polityce
 - Rozprzestrzenianie się choroby, opinii, plotek, mody, ...
- Technika:
 - Sieci energetyczne
 - Sieci transportowe (kolejowe, lotnicze, drogowe, metro), ...
 - Sieci komputerowe, usługi telekomunikacyjne (internet, sieć WWW)
 - Rozprzestrzenianie się awarii, wirusów, ...
- Muzyka:
 - Nowe aranżacje muzyczne na podstawie sieci złożonej reprezentującej utwór muzyczny, ...
- Językoznawstwo, bibliometria:
 - Zależności między wyrazami w książce
 - Zależności między definicjami w słownikach
 - Cytowania publikacji, ...
- Religioznawstwo:
 - Zależności między postaciami w Świętych Księgach (Nowy Testament, Stary Testament, Tora, Koran), ...
- Sport:
 - Analiza gier zespołowych, taktyk, wpływu pojedynczych graczy na grę drużyny i inne.

Na świecie trwają obecnie intensywne prace nad dokładnym badaniem sieci złożonych. Próbuje się znaleźć odpowiedzi na następujące pytania:

- czy istniejące algorytmy w „zwykłych” sieciach są tak samo skuteczne, wydajne, dokładne w sieciach złożonych?
- jakie jeszcze obiekty rzeczywiste (oprócz zbadanych) można adekwatnie modelować z wykorzystaniem sieci złożonych (*Scale Free* i *Small World*)?; dla przykładu, czy bardzo popularne ostatnio sieci semantyczne można modelować i analizować z wykorzystaniem podobnych mechanizmów? [28]
- jak modelować przebieg ewolucji sieci oraz adekwatność opracowanego modelu?

Biorąc pod uwagę przekazane informacje oraz przeprowadzone analizy, autor ma nadzieję, że Czytelnik nie będzie miał problemu z odpowiedzią na pytanie tytułowe: czy sieci rządzą światem? Jeżeli jednak takie wątpliwości się pojawiają, to należy sięgnąć choćby po znakomitą książkę A. i P. Fronczaków [14].

8. Bibliografia

- [1] R. Antkiewicz, M. Chmielewski, R. Kasprzyk, A. Najgebauer, "The prediction of terrorist threat on basis of semantic associations and complex network evolution", *Military Communications and Information Systems Conference'2007*, Behorden Spiegel, Bonn (2007).
- [2] R. Antkiewicz, A. Najgebauer, J. Rulka, Z. Tarapata, R. Wantoch-Rekowski, "Knowledge-Based Pattern Recognition Method and Tool to Support Mission Planning and Simulation", *ICCCP 2011, Part I, Lecture Notes in Computer Science*, Vol. 6922, Springer-Verlag Berlin Heidelberg, 478–487 (2011).
- [3] A.-L. Barabási, R. Albert, "Emergence of scaling in random networks", *Science*, Vol. 286, 509–512 (1999).
- [4] C. Bartosiak, R. Kasprzyk, Z. Tarapata, "Application of Graphs and Networks Similarity Measures for Analyzing Complex Networks", *Biuletyn Instytutu Systemów Informatycznych*, Nr 7, 1–7 (2011).
- [5] C. Bartosiak, *Analiza modeli i algorytmów ewolucji sieci społecznych na przykładzie wybranej sieci rzeczywistej*, praca magisterska (kierownik: Z. Tarapata), WAT, Warszawa, 2012.
- [6] B. Bollobas, *Modern Graph Theory*, Springer-Verlag, Berlin Heidelberg, 2001.
- [7] U. Brandes, "A Faster Algorithm for Betweenness Centrality", *Journal of Mathematical Sociology*, Vol. 25, 163–177 (2001).
- [8] R. Diestel, *Graph Theory*, Springer-Verlag, Berlin Heidelberg, 2005.
- [9] J. Duch, J. S. Waitzman, L.A.N. Amaral, "Quantifying the Performance of Individual Players in a Team Activity", *PLoS ONE* 5(6): e10937 (2010), doi:10.1371/journal.pone.0010937
- [10] P. Erdős, A. Rényi, "On random graphs I", *Publ. Math. Debrecen*, Vol. 6, 290–297 (1959).
- [11] L. Euler, "Solutio problematis ad geometriam situs pertinentis", *Comment. Academiae Sci. I. Petropolitanae* 8, 128–140 (1736).
- [12] T. Florczak, *Grafowo-sieciowe metody reprezentacji i identyfikacji sytuacji decyzyjnych*, praca magisterska (kierownik: Z. Tarapata), WAT, Warszawa, 2012.
- [13] L. Freeman, "A set of Measures of Centrality Based on Betweenness", *Sociometry*, Vol.40, 35–41(1977).
- [14] A. Fronczak, P. Fronczak, *Świat sieci złożonych: Od fizyki do Internetu*, Wydawnictwo Naukowe PWN, Warszawa, 2009.
- [15] C. Hierholzer, "Ueber die Möglichkeit, einen Linienzug ohne Wiederholung und ohne Unterbrechung zu umfahren", *Mathematische Annalen* 6 (1): 30–32, doi:10.1007/BF01442866 (1873).
- [16] A.S. Ismail, R. Hasni, K.G. Subramanian, "Some applications of Eulerian graphs", *International Journal of Mathematical Science Education*, Technomathematics Research Foundation, Vol. 2, Nr 2, 1 – 10 (2009).
- [17] R. Kasprzyk, "The vaccination against epidemic spreading in complex networks", *Biuletyn Instytutu Systemów Informatycznych*, Vol. 3, 39–43 (2009).
- [18] R. Kasprzyk, B. Lipiński, K. Wilkos, M. Wilkos, C. Bartosiak, "CARE – Creative Application to Remedy Epidemics", *Biuletyn Instytutu Systemów Informatycznych*, Vol. 3, 45–52 (2009).
- [19] R. Kasprzyk, *Modele ewolucji systemów złożonych i metody badania ich charakterystyk dla potrzeb komputerowej identyfikacji potencjalnych sytuacji kryzysowych*, rozprawa doktorska, Wojskowa Akademia Techniczna, Warszawa, 2011.
- [20] V. Krebs, "Mapping Networks of Terrorist Cells", *Connections*, Vol. 24(3), 43–52 (2002).

- [21] J. Leskovec, Horvitz E., "Planetary-Scale Views on an Instant-Messaging Network", *Proceedings of the 17th International World Wide Web Conference (WWW'2008)*, April 21–25, 2008, Beijing, China, (915–924) Beijing, 2008.
- [22] X. Liu, C.K. Tse and M. Small, "Composing music with complex networks", *International Conference on Complex Sciences: Theory and Applications (COMPLEX2009)*, 2196–2205, Shanghai, 2009.
- [23] J. L. Pena, H. Touchette, "A network theory analysis of football strategies", *arXiv:1206.6904v1 [math.CO]* (2012).
- [24] P. Pevzner, H. Tang, M. Waterman, "An Eulerian trail approach to DNA fragment assembly", *Proceedings of the National Academy of Sciences of the United States of America* 98 (17): 9748–9753 (2001).
- [25] F. Radicchi, "Who is the best player ever? A complex network analysis of the history of professional tennis", *arXiv:1101.4028v1 [physics.soc-ph]* (2011).
- [26] Z. Tarapata, "Multicriteria weighted graphs similarity and its application for decision situation pattern matching problem", *Proceedings of the 13th IEEE/IFAC International Conference on Methods and Models in Automation and Robotics MMAR'2007*, August 27–30, Szczecin, Poland, 2007.
- [27] Z. Tarapata, R. Kasprzyk, "An application of multicriteria weighted graph similarity method to social networks analyzing", *Proceedings of the 2009 International Conference on Advances in Social Network Analysis and Mining*, July 20–22, 2009, Athens (Greece), *IEEE Computer Society*, 366–368 (2009).
- [28] Z. Tarapata, M. Chmielewski, R. Kasprzyk, "An Algorithmic Approach To Social Knowledge Processing And Reasoning Based On Graph Representation – A Case Study", *ACIIDS 2010, Lecture Notes in Artificial Intelligence*, 5991, Springer-Verlag Berlin Heidelberg, 93–104 (2010).
- [29] Z. Tarapata, R. Kasprzyk, "Graph-based optimization method for information diffusion and attack durability in networks", *RSCTC 2010, Lecture Notes in Artificial Intelligence*, 6086, Springer-Verlag Berlin Heidelberg, 698–709 (2010).
- [30] Z. Tarapata, *Models and algorithms for knowledge-based decision support and simulation in defence and transport applications*, rozprawa habilitacyjna, Wojskowa Akademia Techniczna, Warszawa, 2011.
- [31] Z. Tarapata, *Czy sieci rządzą światem? – Od Eulera do Barabasiego*, wykład inauguracyjny, WAT, Warszawa, 2 października 2012 r.
- [32] J. Travers, S. Milgram, "An Experimental Study of the Small World Problem", *Sociometry*, Vol. 32, Nr 4, 425–443 (1969).
- [33] D. J. Watts, S. Strogatz, "Collective dynamics of 'small-world' networks", *Nature*, Vol. 393, 440–442 (1998).
- [34] B.S. Yilmaz, O. D. Gunel, "Probable applications of complex networks in public relations practices: A scenario-based approach", *PRism* 6(1) (2009), http://praxis.massey.ac.nz/prism_online_journ.html.
- [35] History of Mathematics: On Leonhard Euler (1707-1783), *ScienceWeek* (2003), 6 Nov. 2005, <http://scienceweek.com/2003/sc031121-6.htm>.
- [36] <http://vator.tv/news/2010-12-28-top-10-biggest-social-media-stories-of-2010>
- [37] http://www.fernfachhochschule.ch/ffhs/afe/lws/forschung/research/view?set_language=en
- [38] http://www.beyondthemap.ca/english/explore_web.html
- [39] <http://about-france.com/paris-metro-map.htm>
- [40] <http://www.sciencedirect.com/science/article/pii/S0375960108015867>
- [41] http://da.wikipedia.org/wiki/Fil:High_Speed_Railroad_Map_Europe_2011.png
- [42] <http://www.cut.nazwa.pl/ptg/konfer/include/wydaw.php?r=2007&ab=7b>
- [43] <http://www.mslima.com/mfadt/thesis/2004/08/transportation-routes.html>
- [44] <http://www.vectorstock.com/royalty-free-vector/circulation-of-blood-vector-97672>
- [45] http://www.alz.org/brain_polish/05.asp
- [46] http://www.asbmb.org/asbmbtoday/asbmbtoday_article.aspx?id=14931
- [47] <http://phys.strath.ac.uk/information/acadstaff/ernesto.estrada.php>
- [48] <http://www.crossway.org/blog/2007/01/mapping-nt-social-networks/>
- [49] <http://www.dailymail.co.uk/sciencetech/article-2064746/Facebook-shrinks-degrees-separation-just-FOUR.html>
- [50] <http://thehypertextual.com/2011/03/07/middle-east-strategy-lesson-social-networks-beat-weapons/>

- [51] http://strona.wcy.wat.edu.pl/index.php?option=com_content&task=view&id=247&Itemid=131
- [52] <http://www.youtube.com/watch?v=2guKJfvq4uI>

Do Networks Control the World? From Euler to Barabasi

Z. TARAPATA

In the paper problem of multi-applications of graph and networks theory models and methods is considered. The paper is especially focused on specific features of real networks. Analysis methods of these networks, their properties and what result from these analyses are discussed. Algorithmic aspects focused on computational complexity of methods to compute values of network's characteristics are considered. Several interdisciplinary examples of real networks illustrate network's properties being described. In conclusions current and future trends in development of complex networks are considered.

Keywords: complex networks, complex systems, scale free, small world, applications of complex networks, algorithmic aspects of complex networks.