

Modelowanie i analiza sieci złożonych

II. Historia *sieciologii*. Przykłady sieci rzeczywistych.

Grzegorz Siudem

Politechnika Warszawska



**Politechnika
Warszawska**

Unia Europejska
Europejski Fundusz Społeczny



Zadanie 10 pn.

„Przygotowanie i uruchomienie nowego kierunku studiów na studiach II stopnia
- Inżynieria i Analiza Danych (IAD)”

realizowane jest w ramach projektu
„NERW PW. Nauka – Edukacja – Rozwój – Współpraca”
współfinansowanego ze środków Unii Europejskiej
w ramach Europejskiego Funduszu Społecznego

Przed zajęciami

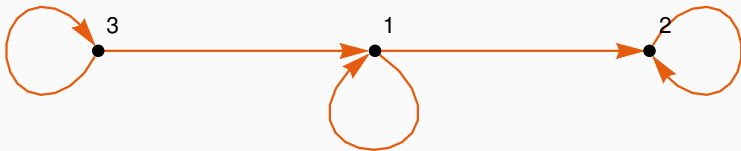
Przypomnij sobie – metody reprezentacji grafów

Macierz sąsiedztwa

$$A = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix}.$$

Lista sąsiedztwa

$$L = \{\{1,2\}, \{2\}, \{1,3\}\}.$$



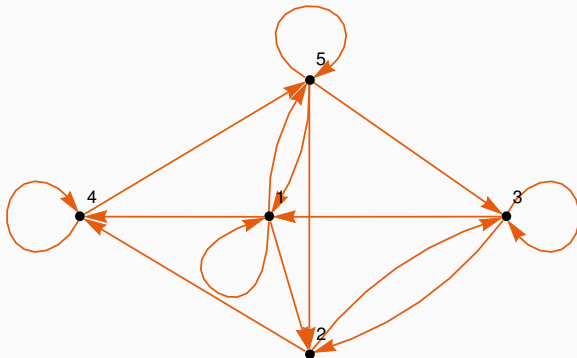
Przypomnij sobie – stopień wierzchołka

Stopień wierzchołka

Liczba wchodzących albo wychodzących do wierzchołka krawędzi.

$$k_{out} = \{4, 2, 3, 2, 4\},$$

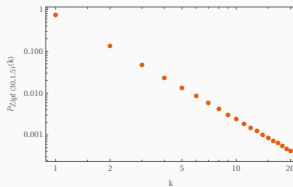
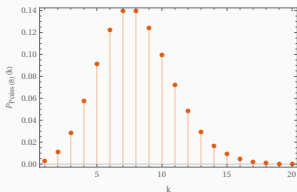
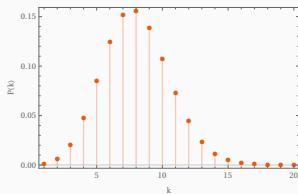
$$k_{in} = \{3, 3, 3, 3, 3\}.$$



Przypomnij sobie – dyskretne rozkłady prawdopodobieństwa

Dyskretne rozkłady prawdopodobieństwa

- rozkład dwumianowy,
- rozkład Poissona,
- rozkład Zipfa.



$$P(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$$P(k) = \frac{\lambda^k e^{-\lambda}}{k!}$$

$$P(k) = \frac{1/k^s}{H_{N,s}}$$

Przypomnienie

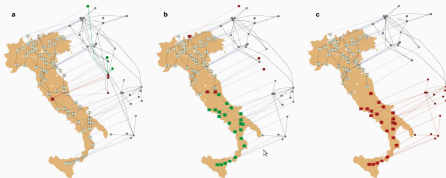
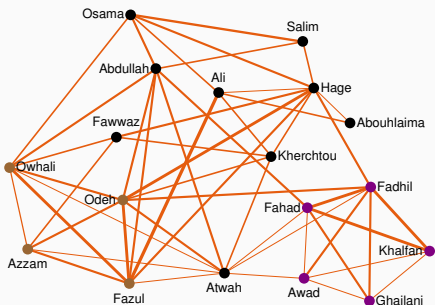
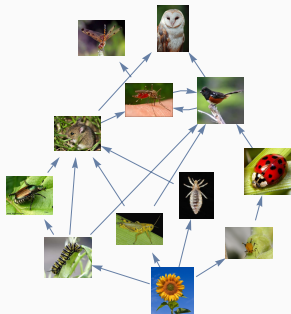
W *sociologii* często zastępuje się rozkłady dyskretne ciągłymi (o tym dlaczego porozmawiamy na zajęciach).
Znajdź ciągłe analogi rozkładów z poprzedniego slajdu.

Wykład

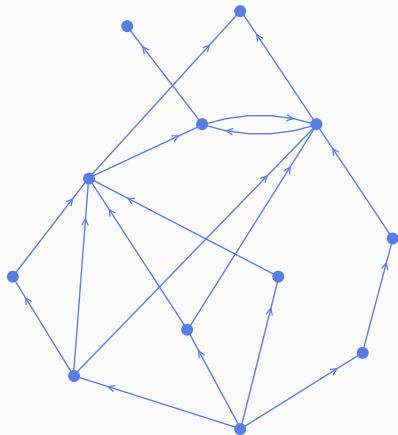
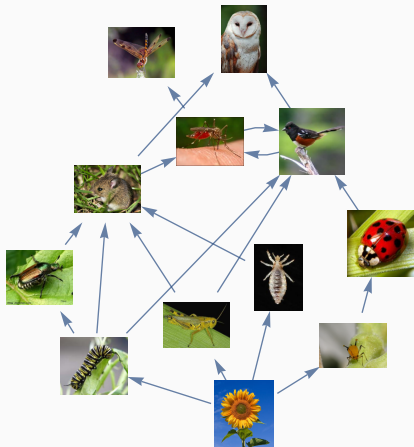
Dlaczego sieci/grafy są tak użyteczne?

Grafy są ilustracją relacji

Jakie relacje ilustrują poniższe grafy?



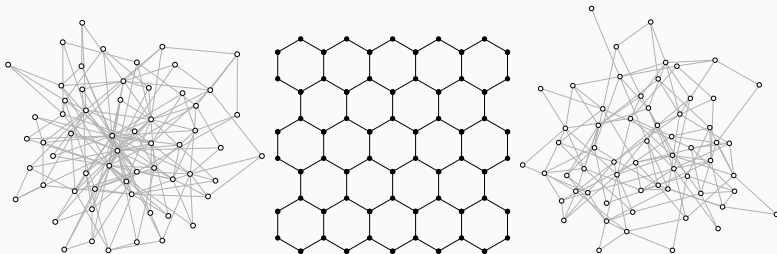
Czym się różnią sieci od grafów?



Dlaczego sieci są złożone?

Cechy rzeczywistych sieci

- rozkłady stopni wierzchołków mają grube ogony.
- wierzchołki są heterogeniczne.
- występują zjawiska małych światów (sześć kroków?).
- wierzchołki są ze sobą skorelowane (współczynnik gronowania).



Który z tych grafów przedstawia sieć rzeczywistą?

Co to jest *sieciologia*?

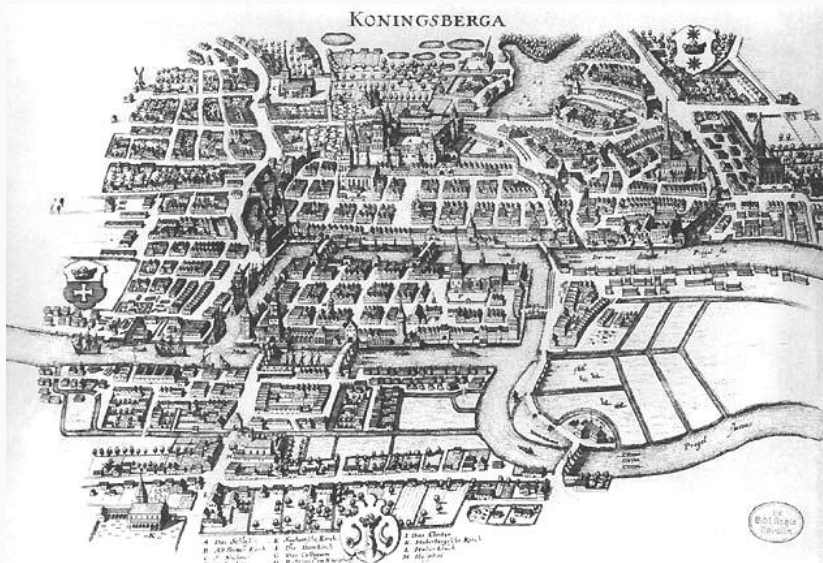
Ang. *Network Science* to połączenie narzędzi

- matematyki (teoria grafów, probabilistyka, statystyka),
- fizyki (narzędzia fizyki statystycznej, przejścia fazowe),
- informatyki (metody wizualizacji sieci, przechowywanie i przetwarzanie dużych sieci),
- analizy i inżynierii danych,

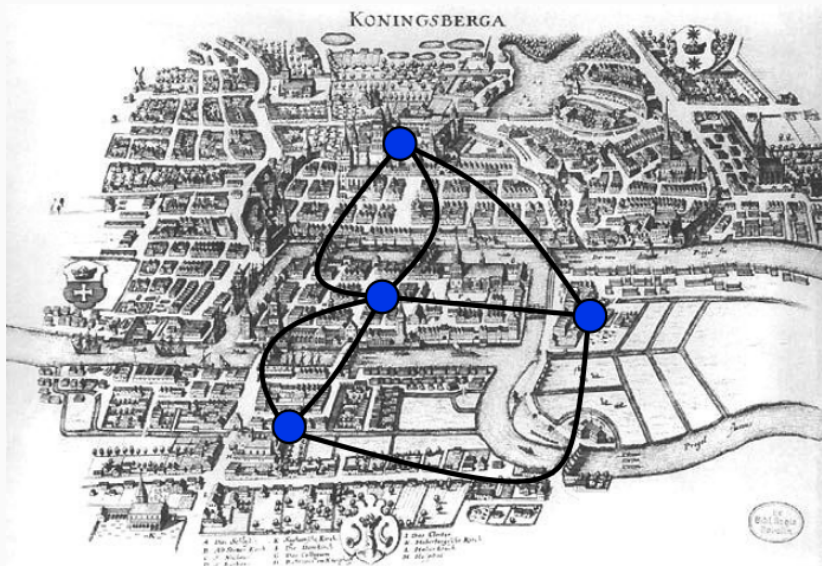
z wiedzą specjalistyczną z

- socjologii,
- ekonomii,
- biologii,
- medycyny,
- inżynierii,
- i wielu innych...

Historia sieciologii – problem mostów królewskich (1736)



Historia sieciologii – problem mostów królewskich (1736)



Duże zainteresowanie socjologów badaniami sieci społecznych

Najbardziej cytowana praca z tej tematyki:

M. S. Granovetter. The strength of weak ties. American Journal of Sociology, 78: 1360, 1973.

- badania głównie jakościowe,
- niezależne od badań matematyków z teorii grafów,
- nie skupiamy się na tym kierunku badań (poza jednym wyjątkiem!).

Pierwszy ilościowy model sieci przypadkowych

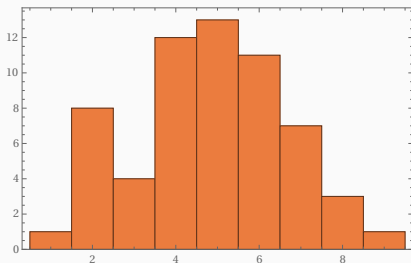
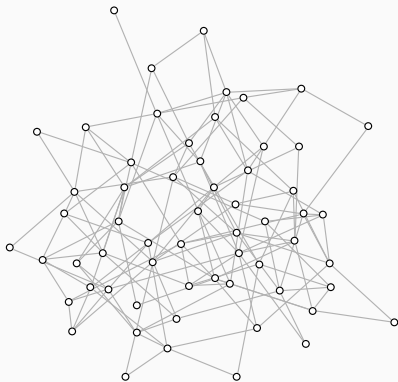
P. Erdős, A. Rényi. On random graphs. Publicationes Mathematicae, 6: 290, 1959.

- prosty model probabilistyczny w którym łączymy przypadkowo wierzchołki krawędziami.
- występuje w dwóch wariantach:
 - Rozkładamy E krawędzi pomiędzy N wierzchołków.
 - Każdą parę spośród N wierzchołków łączymy krawędzią z prawdopodobieństwem p .

Więcej szczegółów na wykładzie 5.

Rozkład stopni wierzchołków

$P(k) =$ ułamek węzłów ze stopniem k



Czy to jest rzeczywista sieć?

Uwaga!

Nie należy mylić eksperymentu *korespondencyjnego* ze słynnym eksperymentem posłuszeństwa względem autorytetów!

Słynne sześć uścisków dłoni

Milgram, Stanley, The Small World Problem, Psychology Today, 1967.

Opis eksperymentu

- Milgram rozsyła do 160 przypadkowo wybranych ludzi list z opisem eksperymentu.
- Celem doświadczenia było dostarczenie listu do pewnego, przypadkowego adresata, którego zdjęcie, adres i dane umieszczone były w liście.
- Jeśli odbiorca listu znał adresata osobiście powinien mu go dostarczyć, w przeciwnym wypadku powinien go wysłać do kolejnej osoby, którą podejrzewał, że może znać adresata.

Wnioski

- Większość listów zaginęła.
- Do celu dotarło tylko 42 ze 160 wysłanych.
- Niektóre potrzebowały tylko dwóch pośredników by dotrzeć do celu!
- Średnio, listy, które dotarły przekazywane były dalej tylko sześć razy.

Ciekawostka:

- Liczbą Erdösa naukowca nazywamy liczbę prac, jaka dzieli go od Paula Erdösa.
- Liczbą Bacona aktora nazywamy liczbę filmów jaka dzieli go od Kevina Bacona.

Problem:

Grafiy poissonowskie nie mają cechy małych światów!

Model WS

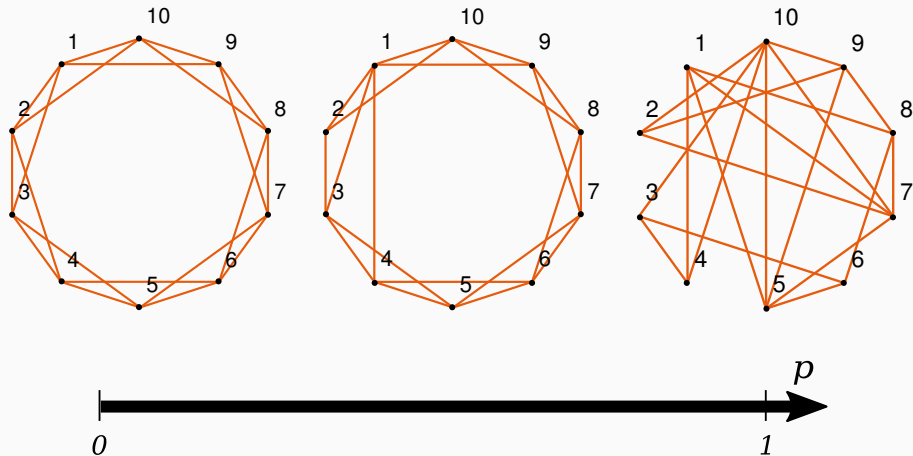
D.J. Watts, S.H. Strogatz, Collective dynamics of 'small-world' networks, Nature. 393 440–442 (1998).

Opis:

- Zaczynamy od udekorowanej sieci regularnej z periodycznymi warunkami brzegowymi.
- Z prawdopodobieństwem p przetyczamy każdy z wierzchołków.

Więcej informacji na wykładzie 5.

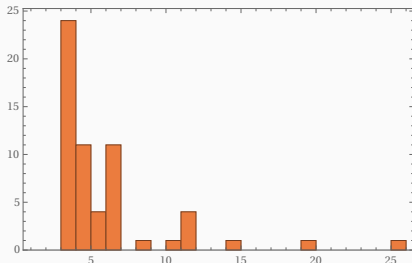
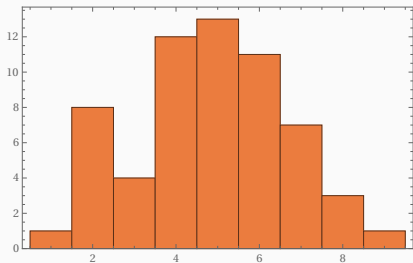
Historia sieciologii – model Watts-Strogatza (1998)



Historia sieciologii – model Barabásiego-Alberty (1999)

Problem:

Sieci rzeczywiste mają potęgowe rozkłady stopni wierzchołków!



Co to oznacza?

- Brak typowej skali.
- grube ogony.
- Epidemie...

Model BA

A.-L. Barabási, R. Albert, *Emergence on scaling in random networks*, *Science*, 286:509-512, 1999.

A.-L. Barabási, R. Albert, H. Jeong, *Mean-field theory for scale-free random networks*, *Physica A*, 272: 173-187, 1999.

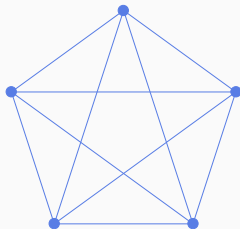
Główne założenia:

- Sieć jest ewoluująca, tzn. z każdym krokiem dodajemy nowy wierzchołek i ją powiększamy.
- Krawędzie dodawane są zgodnie z regułą preferencyjnego dołączania (ang. *rich get richer rule*).

Oba założenia są konieczne!

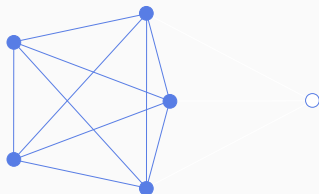
Procedura konstrukcji:

- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.



Procedura konstrukcji:

- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.

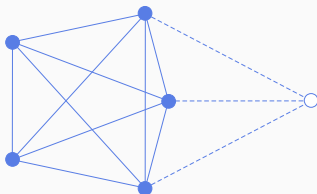


Więcej informacji na wykładzie 5.

Historia sieciologii – model Barabásiego-Alberty (1999)

Procedura konstrukcji:

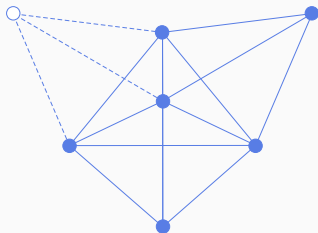
- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.



Więcej informacji na wykładzie 5.

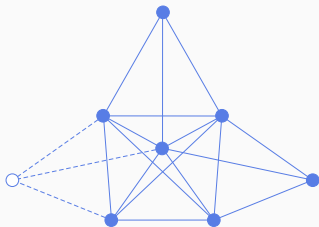
Procedura konstrukcji:

- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.



Procedura konstrukcji:

- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.



Historia sieciologii – model Barabásiego-Alberty (1999)

Procedura konstrukcji:

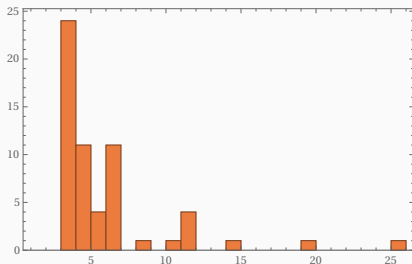
- Zaczynamy od małego grafu pełnego.
- Dodajemy nowy węzeł.
- Dołączamy go krawędziami do istniejących węzłów zgodnie z regułą preferencyjnego dołączania.
- Powtarzamy dwa powyższe punkty.

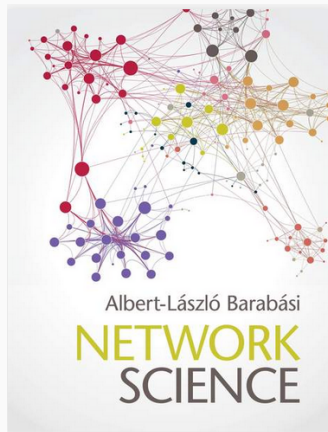


Historia sieciologii – model Barabásiego-Alberty (1999)

Podsumowanie:

- Sieć ewolucyjna o potęgowym rozkładzie stopni wierzchołków.
- Początek nowoczesnej sieciologii.
- Znaczenie tzw. hubów.





Głównym zadaniem nauki o sieciach jest modelowanie rzeczywistych zjawisk i procesów przy pomocy sieci złożonych, będących *de facto* odpowiednimi grafami.

Typowe narzędzia to:

- Ilościowy opis sieci, w tym także
 - Proponowanie metryk sieci (wykład 4).
 - Analiza ich własności (wykłady 4,8).
 - Mierzenie przy ich pomocy sieci rzeczywistych (wykłady 4,8, 13-15).
 - Opis procesów dynamicznych na rzeczywistych sieciach (np. epidemie, propagacja informacji, etc.) (wykłady 8,11-15)
- Teoretyczne modele sieci złożonych (wykłady 5,6,9,10).
- Teoretyczne modele procesów na sieciach (wykłady 11,12).
- Wizualizacja sieci (zarówno teoretycznych jak i rzeczywistych) (wykład 3).

Dziękuję za uwagę!



**Politechnika
Warszawska**

Unia Europejska
Europejski Fundusz Społeczny



Zadanie 10 pn.

„Przygotowanie i uruchomienie nowego kierunku studiów na studiach II stopnia
- Inżynieria i Analiza Danych (IAD)”

realizowane jest w ramach projektu
„NERW PW. Nauka – Edukacja – Rozwój – Współpraca”
współfinansowanego ze środków Unii Europejskiej
w ramach Europejskiego Funduszu Społecznego