

Plan wykładu:

- Motywacja - od Internetu do sieci genetycznych
- Grafy losowe
- Sieci typu małego świata
- Sieci rosnące + preferencyjne przyłączanie
- Podsumowanie

1998 PageRank, Larry Page, Sergey Brin, **Google**

1999 BA model, Albert-László Barabási, Réka Albert

Emergence of scaling in random networks

[AL Barabási](#), [R Albert](#) - science, 1999 - science.org

Systems as diverse as genetic networks or the World Wide Web are best described as networks with complex topology. A common property of many large networks is that the vertex ...

☆ Save  Cite Cited by 41974 Related articles All 68 versions

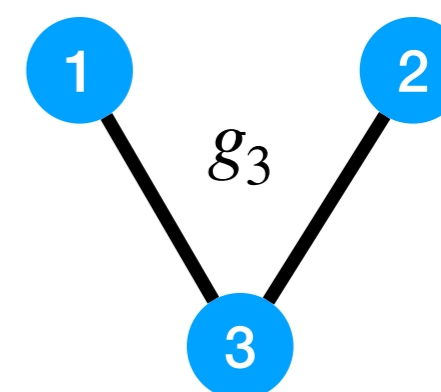
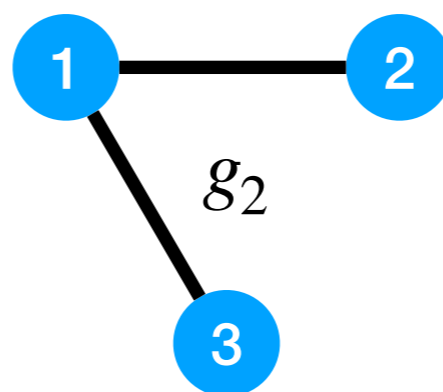
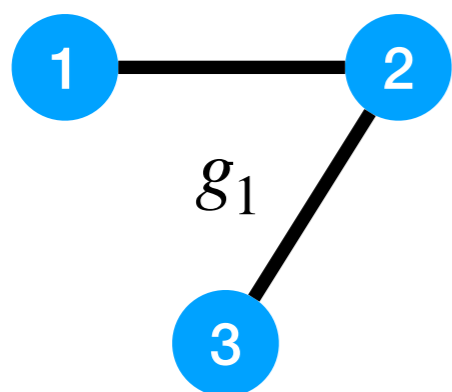
- internet, www
- sieci transportowe
- sieci ekologiczne
- sieci społeczne
- sieci abstrakcyjne (np. regulacja genów, sieci metaboliczne, itp)

Model Erdösa-Rényi'ego-Gilberta 1959 (Paul Erdős, Alfréd Rényi, Edgar Gilbert)

Model $G(n, m)$ = grafy losowe mające n wierzchołków i m krawędzi;
wszystkie grafy są równie prawdopodobne;

Model $G(n, p)$ = grafy losowe mające n wierzchołków;
dowolne dwa wierzchołki połączone są krawędzią
z prawdopodobieństwem p ;

Przykład: $G(3,2)$



$$p(g_1) = p(g_2) = p(g_3) = \frac{1}{3}$$

Ile jest grafów G(n,m)?

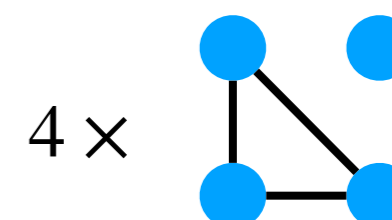
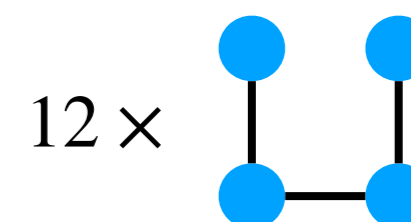
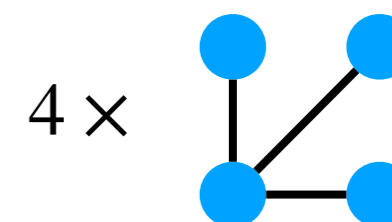
Liczba par wierzchołków: $N = \binom{n}{2} = \frac{n(n-1)}{2}$

Liczba grafów: $|G(n, m)| = \binom{N}{m}$

Przykład: $n = 4; m = 3; N = 6;$

$$|G(4,3)| = \binom{6}{3} = 20$$

Prawdopodobieństwo grafu: $p(g_i) = \frac{1}{20} \quad \forall i = 1, \dots, 20$



Liczba grafów: $|G(n, p)| = 2^N = 2^{\binom{n}{2}}$

Prawdopodobieństwo grafu: $p^m(1 - p)^{N-m}$

Prawdopodobieństwo, że graf ma m krawędzi: $p_E(m) = \binom{N}{m} p^m(1 - p)^{N-m}$

Średnia liczba krawędzi: $\langle m \rangle = \sum_{m=0}^N m p_E(m) = Np$

Standardowe odchylenie: $\sigma(m) = \sqrt{Np(1 - p)}$

$$\frac{\sigma(m)}{\langle m \rangle} = \frac{1}{\sqrt{N}} \sqrt{\frac{1-p}{p}} \rightarrow 0, \text{ gdy } Np \rightarrow \infty$$

W granicy $n^2p \rightarrow 0$, grafy $G(n, p)$ zachowują się jak $G(n, m)$, dla $m = Np$

Grafy $G(n, p)$ są łatwiejsze do analizy niż $G(n, m)$, bo krawędzie są niezależne

Stopień wierzchołka: k_i = liczba krawędzi wychodzących z wierzchołka i

Lemat u uściskach dłoni: $\sum_{i=1}^n k_i = 2m$

Średni stopień: $\langle \bar{k} \rangle = \frac{2\langle m \rangle}{n} = \frac{2Np}{n} = (n-1)p \approx np$

Grafy gęste: $p = \text{const} > 0 \implies \bar{k} \rightarrow \infty$

Grafy rzadkie: $p = \frac{\alpha}{n}$; $\alpha = \text{const} > 0 \implies k \rightarrow \alpha$

Rzadkie grafy losowe: $G\left(n, p = \frac{\alpha}{n}\right)$, $n \rightarrow \infty$; $\alpha = \text{const} > 0$

Asymptotycznie w granicy $n \rightarrow \infty$:

$\alpha < 1$: graf składa się z wielu drobnych spójnych składowych o maksymalnym rozmiarze $\sim O(\log n)$

$\alpha = 1$: prawie na pewno pojawia się składowa o rozmiarze $\sim O(n^{2/3})$

$\alpha > 1$: prawie na pewno pojawia się gigantyczna składowa o rozmiarze $\sim O(n)$

Dla $p > \frac{c \log n}{n}$ i $c > 1$ gigantyczna składowa prawie na pewno ma rozmiar n

Prawdopodobieństwo, że wierzchołek ma stopień k :

$$p(k) = \text{Prob}(k_i = k) = \binom{n-1}{k} p^k (1-p)^{n-k-1} \quad \text{rozkład dwumianowy}$$

$$\text{Dla } p = \frac{\alpha}{n} \quad \text{i } n \rightarrow \infty \quad \Longrightarrow \quad p(k) = \frac{\alpha^k}{k!} e^{-\alpha} \quad \text{rozkład Poisson'a}$$

$$\langle k \rangle = \alpha; \quad \sigma(k) = \sqrt{\alpha}$$

Prawdopodobieństwo, że wystąpi duże k jest bardzo małe!

Grafy $G(n, p)$ są często używane jako punkt odniesienia (benchmark), dla innych grafów losowych i sieci rzeczywistych;

Dlatego bada się statystyki występowania różnych motywów (podgrafów) na tych grafach.

Najprostszym takim motywem jest cykl trójkątny?

Pytanie: ile jest cykli trójkątnych na grafach $G\left(n, p = \frac{\alpha}{n}\right)$ dla $n \rightarrow \infty$

Średnia liczba trójkątów: $\langle T \rangle = \binom{n}{3} p^3 \approx \frac{\alpha^3}{6}$ (stała niezależna od n)

np. dla $\alpha = 4$; $\langle T \rangle \approx 10.67$

Podobnie można pokazać, że jest skończona liczba cykli o długości 4,5,...

Współczynnik klastrowania (gronowania)

Współczynnik klastrowania = prawdopodobieństwo, że sąsiedzi wierzchołka są swoimi sąsiadami (tzn. są połączeni bezpośrednią krawędzią):

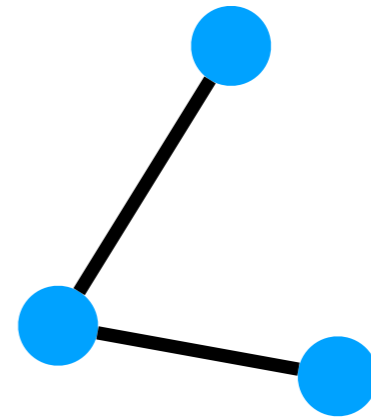
Dla wierzchołka i : $C_i = \frac{2T_i}{k_i(k_i - 1)}$; T_i = liczba trójkątów zawierających i

Jeśli wszystkie sąsiednie wierzchołki są połączeni ze sobą: $T_i = \frac{k_i(k_i - 1)}{2} \Rightarrow C_i = 1$

Średni współczynnik $\bar{C} = \frac{1}{n} \sum_i C_i$

Globalny współczynnik $C_g = \frac{\sum_i T_i}{\sum_i k_i(k_i - 1)/2} = \frac{3T}{Z}$

gdzie Z to liczba zawiasów; zawias to połączona trójka:



Dla grafów Erdősa-Rényi'ego: $\bar{C} \sim C_g \sim \frac{1}{n} \rightarrow 0$

Odległość l_{ij} między wierzchołkami ij , to liczba krawędzi najkrótszej drogi łączącej te wierzchołki (na grafie niespójnym $l_{ij} = \infty$)

$$\text{Średnia odległość: } \bar{l} = \frac{2}{n(n-1)} \sum_{i < j} l_{ij}$$

Średnica grafu: $L = \max(l_{ij})$

Dla spójnej składowej grafu Erdösa-Rényi'ego:

$\bar{l} \sim L \sim \log n$; **efekt małego świata**;

Dla porównania: dla sieci regularnej w d -wymiarach: $\bar{l} \sim L \sim n^{1/d}$

Koniec XX wieku i początek XXI wieku:
statystyczne własności sieci rzeczywistych

- internet, www
- sieci transportowe
- sieci ekologiczne
- sieci społeczne
- sieci abstrakcyjne (np. regulacja genów, sieci metaboliczne, itp)

Większość z tych sieci wykazywała

- ciężko-ogonowe rozkłady stopni - np. $p_k \sim k^{-\gamma}$
występują wierzchołki mające stopnie znacznie większe niż średni stopień
- małe światy
- duży współczynnik klastrowania

Duncan J. Watts, Steven Strogatz, 1998

n wierzchołków, każdy początkowo połączony

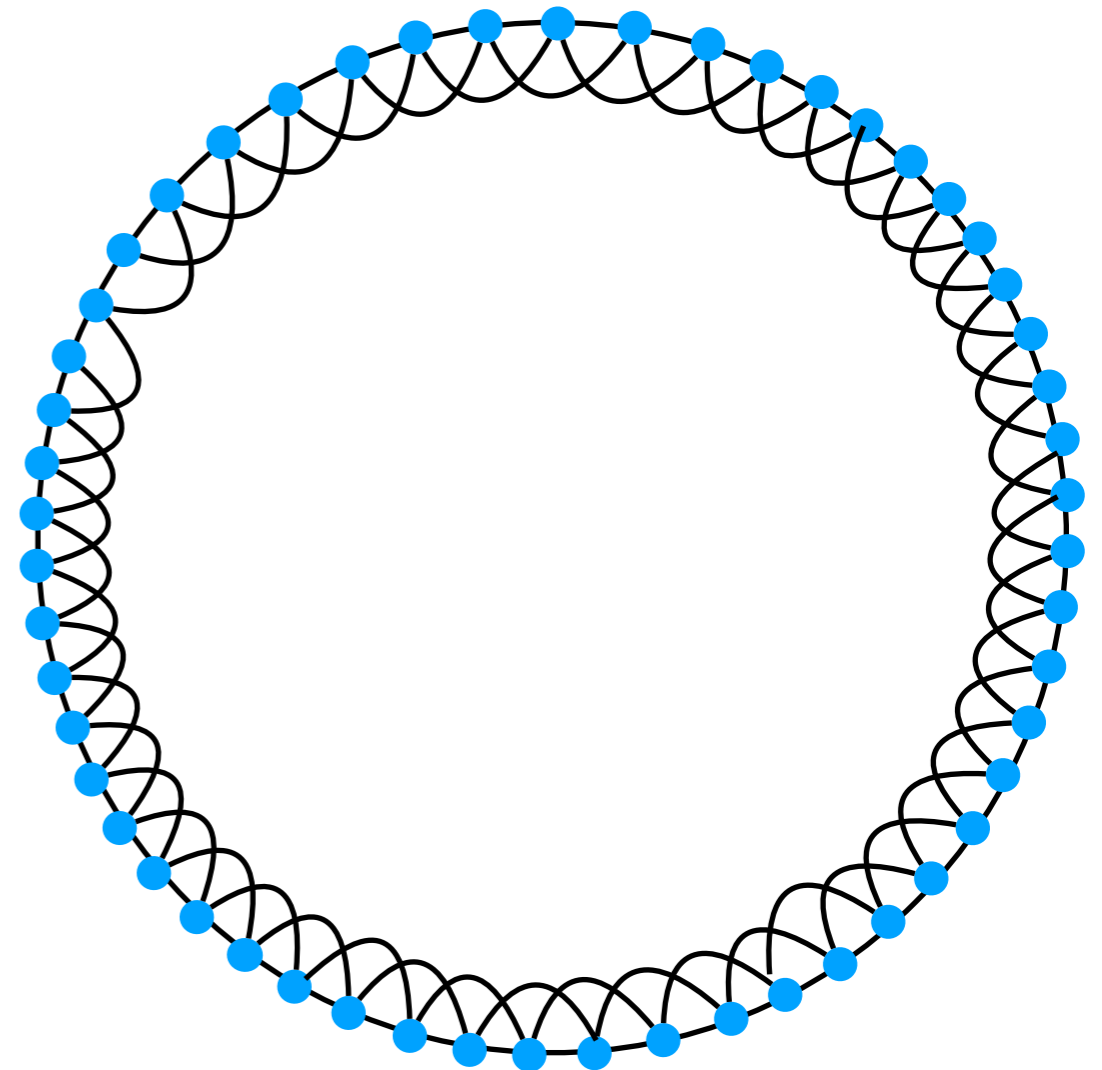
z k sąsiadami (na rysunku z $k = 4$)

Na początku

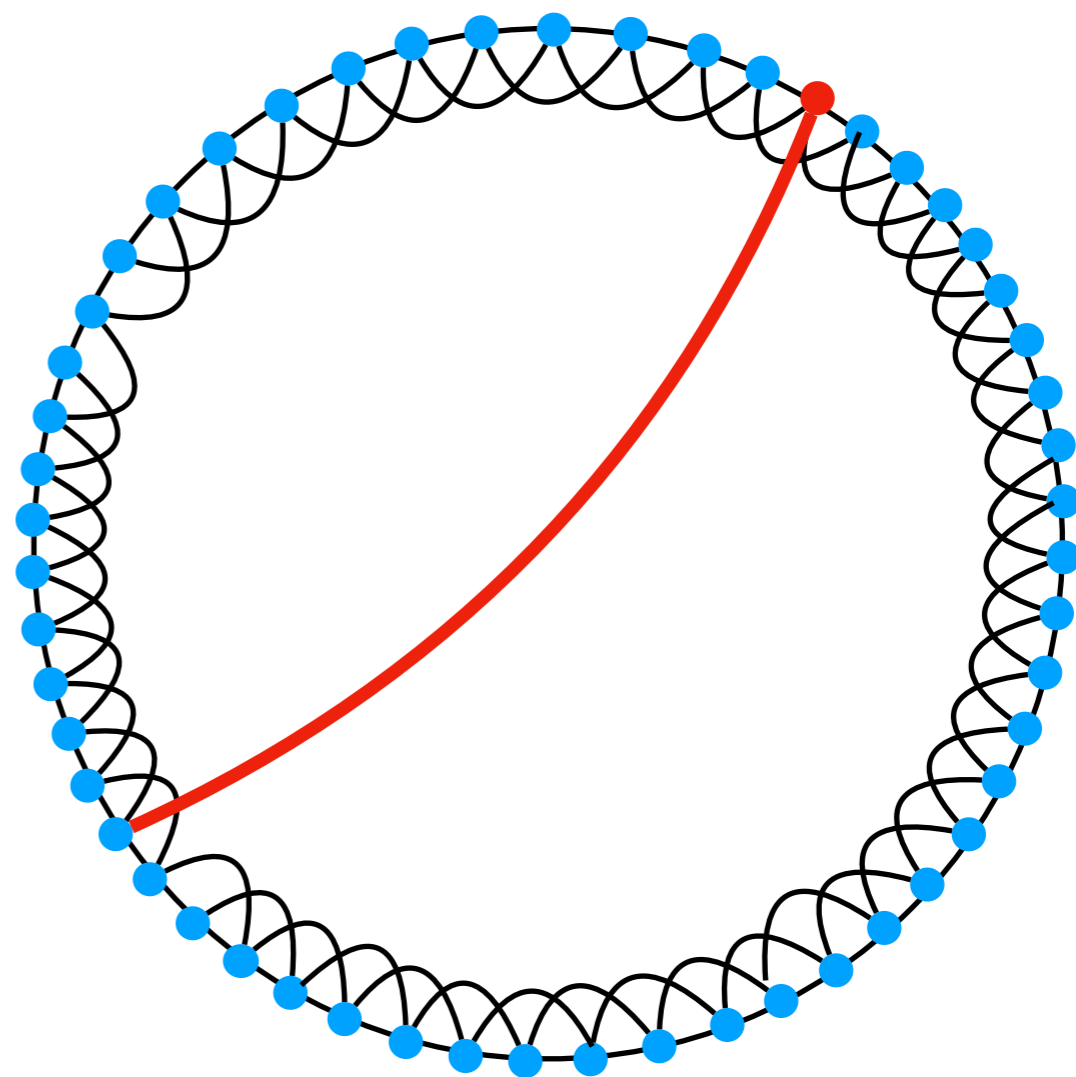
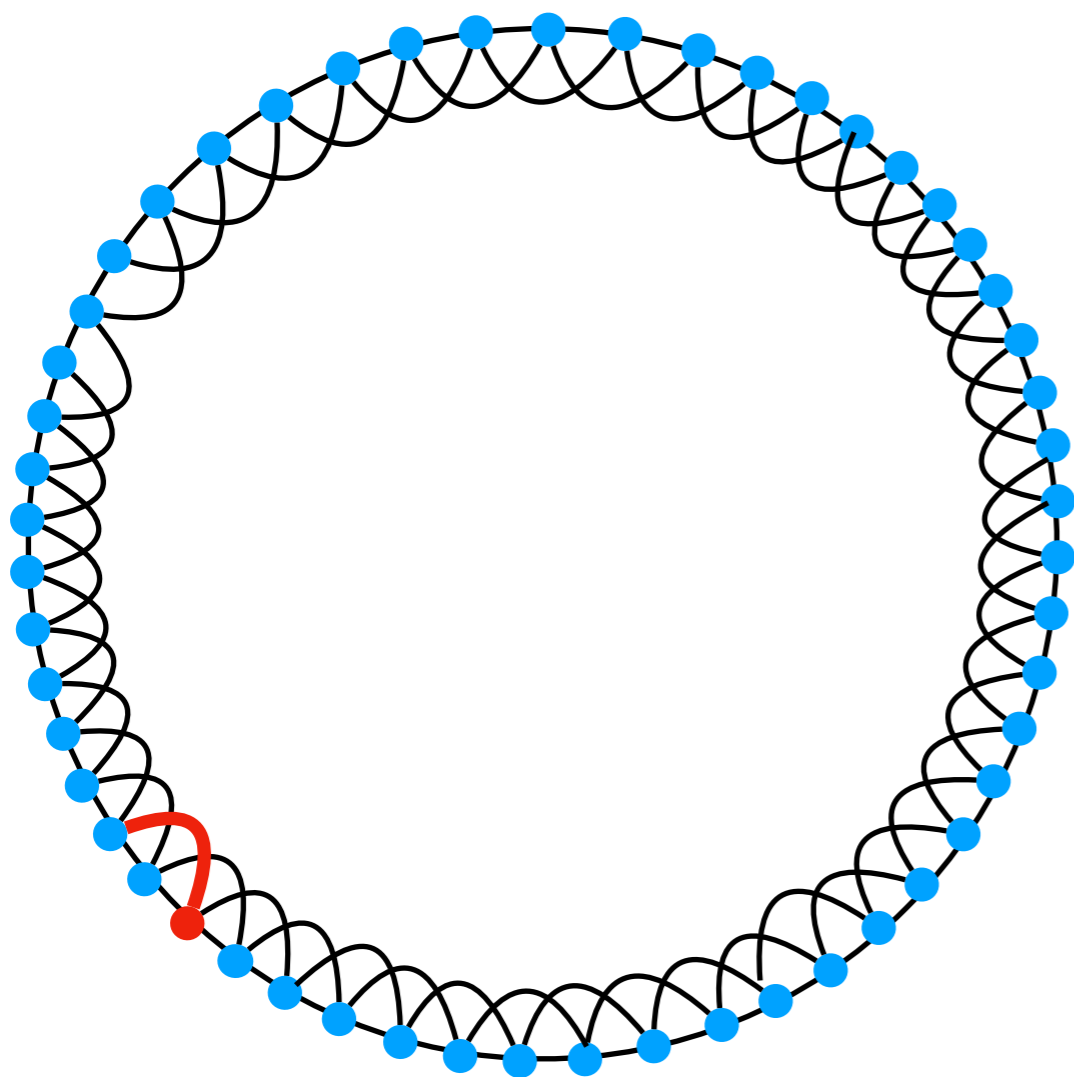
$$\bar{C}_0 = \frac{3(k-2)}{4(k-1)}$$

$$\bar{l}_0 \approx \frac{n}{2k}$$

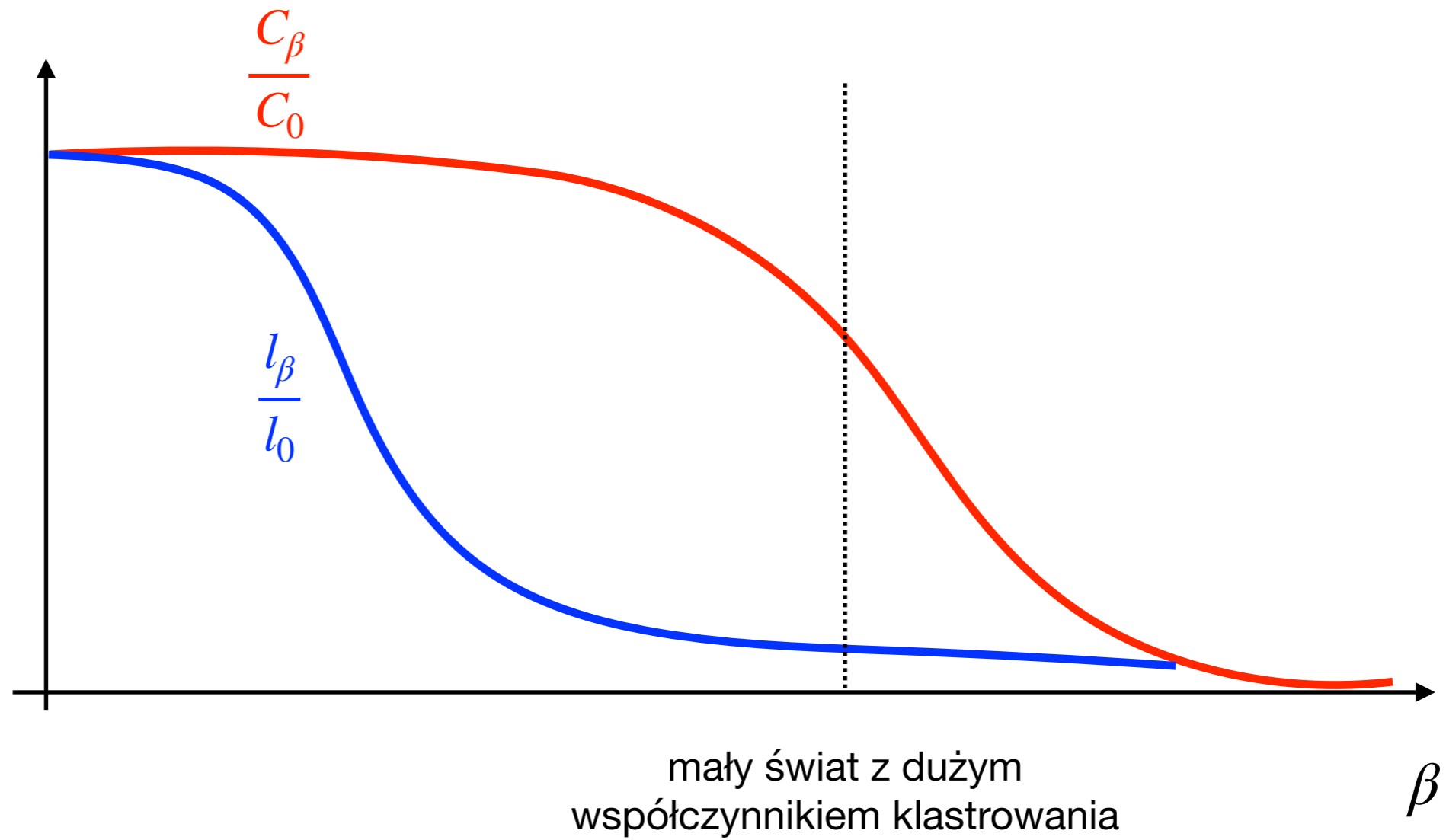
Losowe przepięcia krawędzi (krawędź zmienia losowo jeden wierzchołek końcowy)



Losowe przepięcia krawędzi



$\beta = \frac{E}{nk}$; gdzie E jest liczbą losowo przedrutowanych krawędzie



Alberta László Barabási, Réka Albert, 1999

Sieci rosnące z preferencyjnym przyłączaniem jako model internetu:

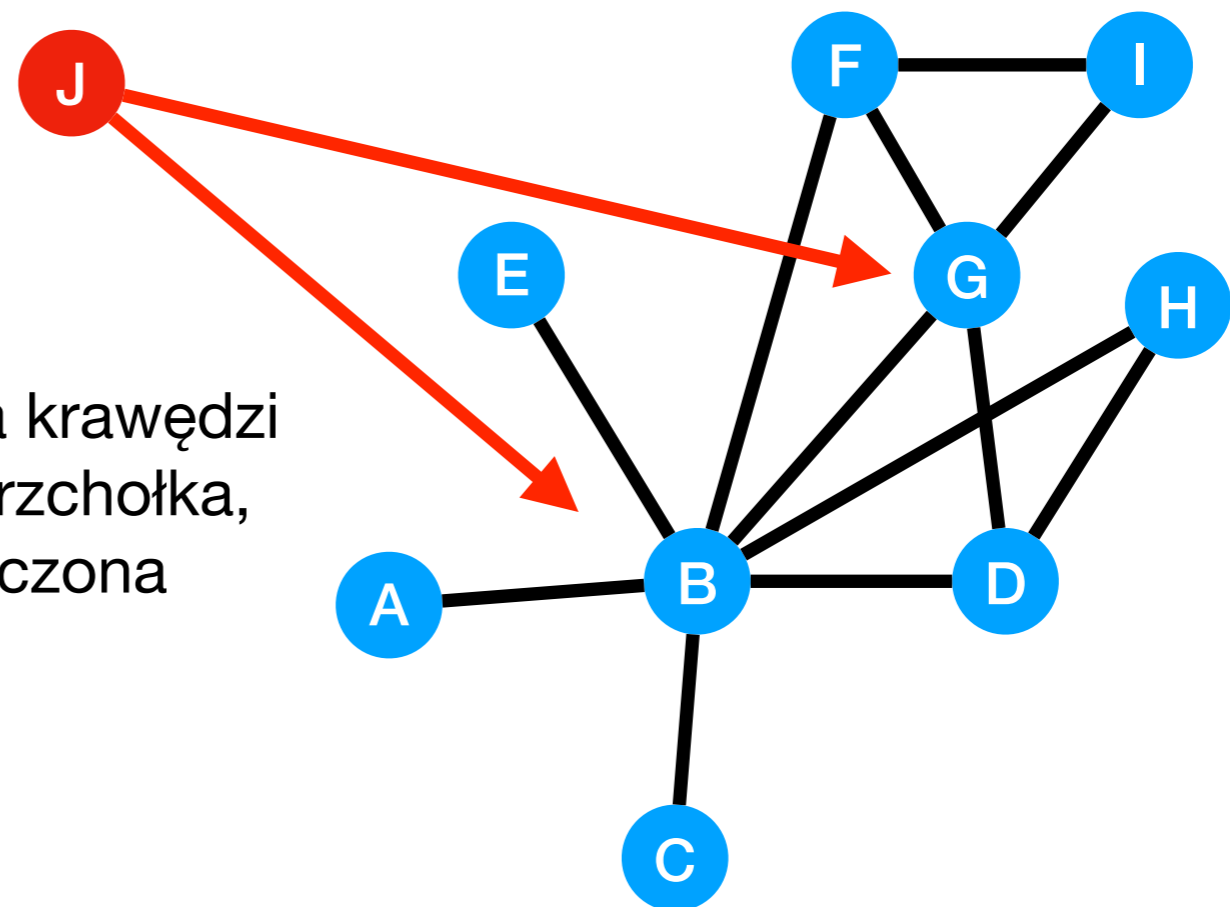
Rozkład potęgowy:

G.U. Yule (1924), Herbert Simon (1955), Derek J. de Solla Price (1965)

Przykład:

prawdopodobieństwo przyłączenia krawędzi jest proporcjonalne do stopnia wierzchołka, do którego krawędź ma być przyłączona

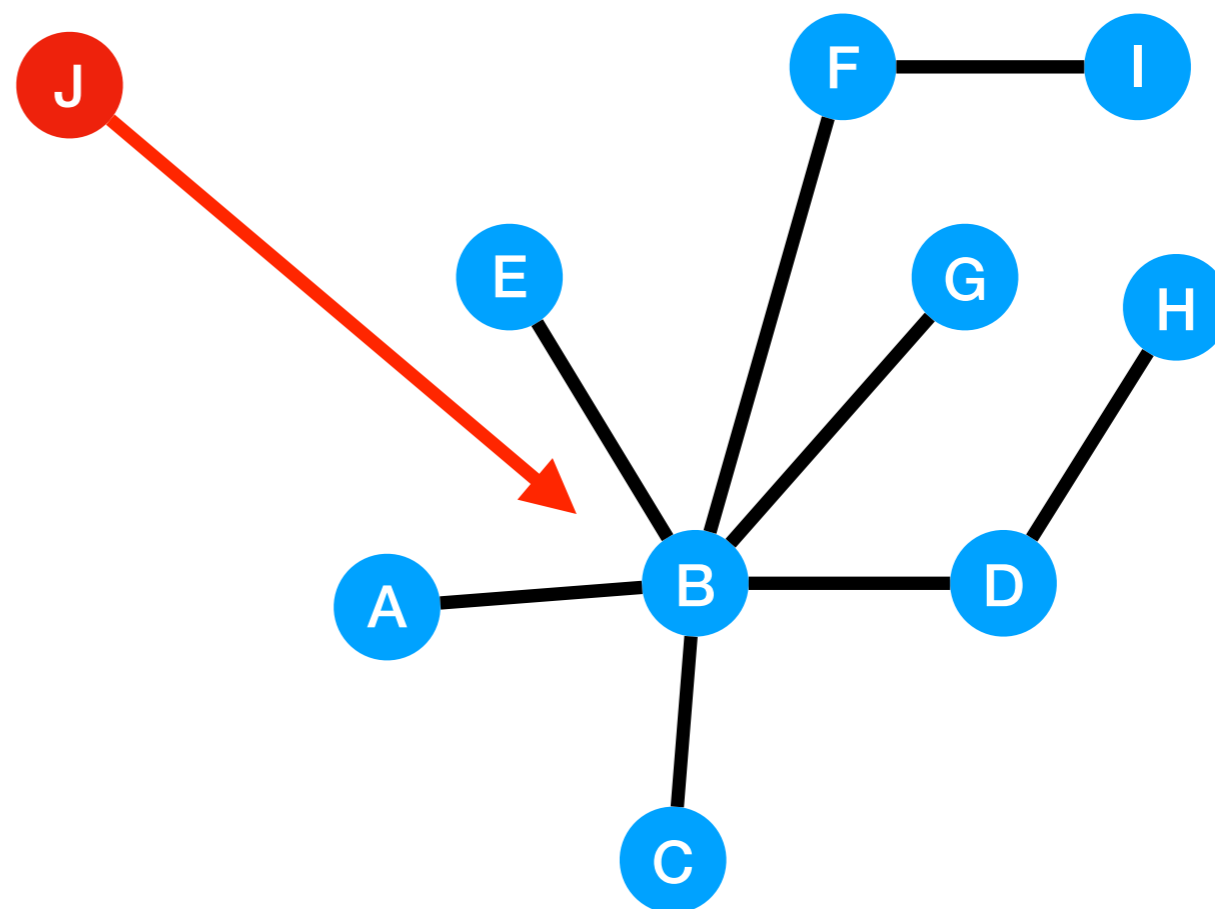
$$\pi_k \sim k$$



Równanie wzrostu: n_k liczba wierzchołków stopnia k ;

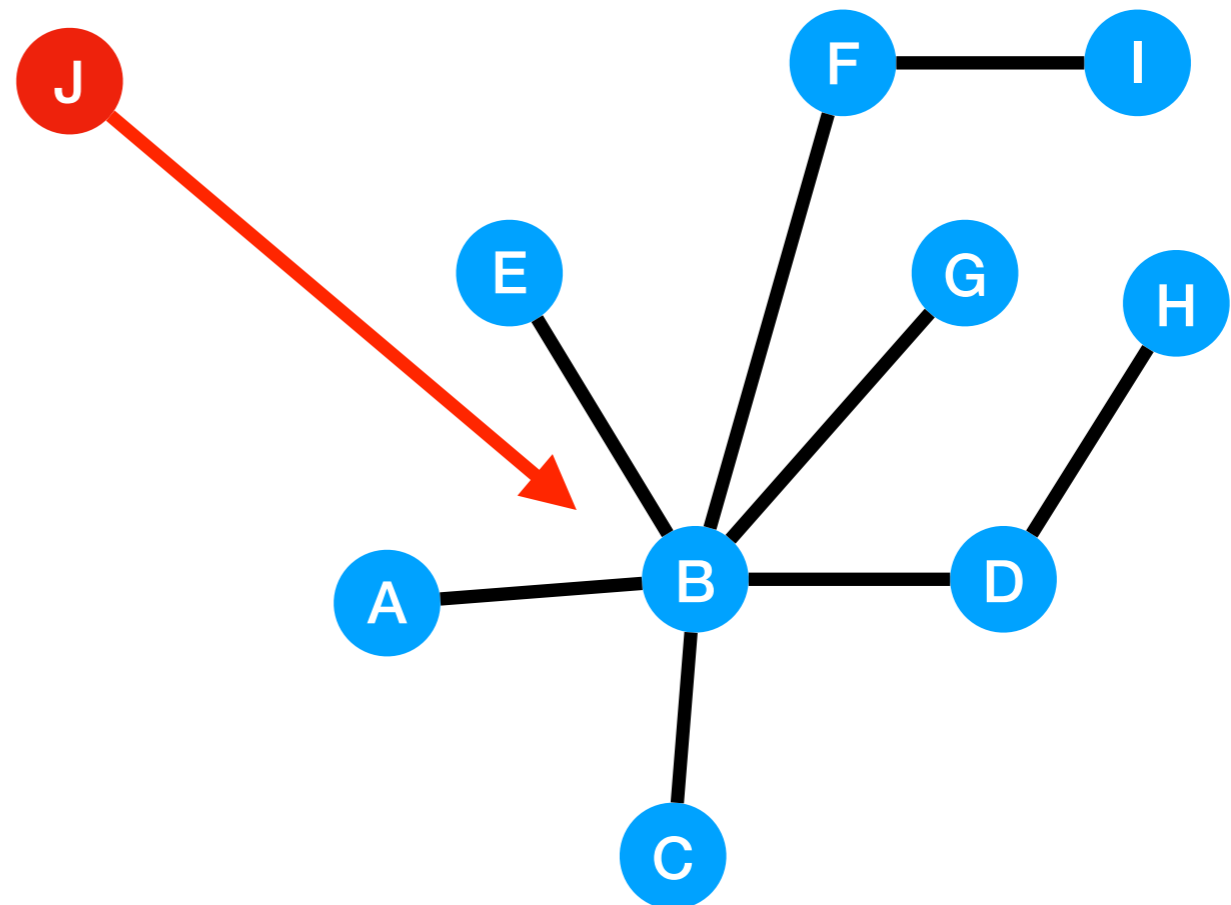
π_k - prawdopodobieństwo przyłączenia do wierzchołka o stopniu k

$$\dot{n}_k = \pi_{k-1}n_{k-1} - \pi_k n_k + \delta_{k,1}$$



Losowe: $\pi_k = \frac{1}{n}$

Preferencyjne: $\pi_k = \frac{k}{\sum_i k_i} = \frac{k}{2m} = \frac{k}{2(n-1)} \approx \frac{k}{2n}$



$$\dot{n}_k = \pi_{k-1}n_{k-1} - \pi_k n_k + \delta_{k,1}$$

$$\dot{n}_k = \frac{n_{k-1}}{n} - \frac{n_k}{n} + \delta_{k,1} = p_{k-1} - p_k + \delta_{k,1}$$

Stan stacjonarny: $\dot{n}_k = p_k$, bo n odgrywa rolę czasu

$$p_k = p_{k-1} - p_k + \delta_{k,1} \implies 2p_k = p_{k-1} + \delta_{k,1}$$

$$p_1 = \frac{1}{2}; \quad p_k = \frac{1}{2}p_{k-1}$$

Rozwiązanie:

$$p_k = \left(\frac{1}{2}\right)^k = e^{-(\log 2)k} \quad \text{dla } k = 1, 2, 3, \dots$$

$$\dot{n}_k = \pi_{k-1}n_{k-1} - \pi_k n_k + \delta_{k,1}$$

$$\dot{n}_k = \frac{(k-1)n_{k-1}}{2n} - \frac{kn_k}{2n} + \delta_{k,1} = (k-1)p_{k-1} - kp_k + \delta_{k,1}$$

Stan stacjonarny: $\dot{n}_k = p_k$,

$$2p_k = (k-1)p_{k-1} - kp_k + 2\delta_{k,1} \implies (k+2)p_k = (k-1)p_{k-1} + 2\delta_{k,1}$$

$$p_1 = \frac{2}{3}; \quad p_k = \frac{k-1}{k+2} p_{k-1}$$

Rozwiązanie:

$$p_k = \frac{4}{k(k+1)(k+2)} \sim k^{-3} \quad \text{dla } k = 1, 2, 3, \dots$$

Liniowe jądro przyłączenia

$$\pi_k \sim k + \omega \implies p_k \sim k^{-\gamma}, \text{ gdzie } \gamma = 3 + \omega; \quad \omega > -1;$$

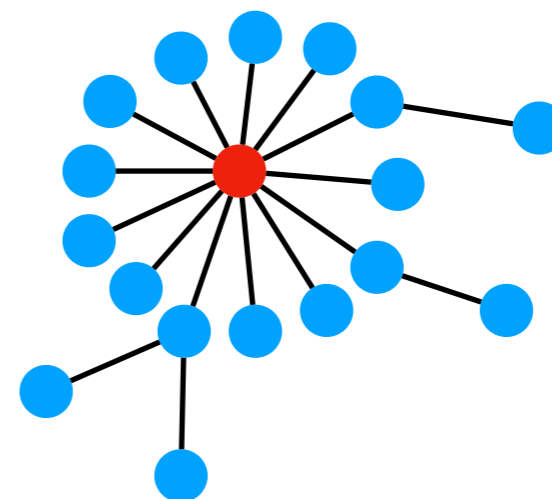
Dla $\gamma \in (-2, 3]$ wariancja rozkładu jest nieskończona !

To prowadzi do pojawiania się dużych wartości k (tzw. huby)

Ponadliniowe jądro przyłączenia

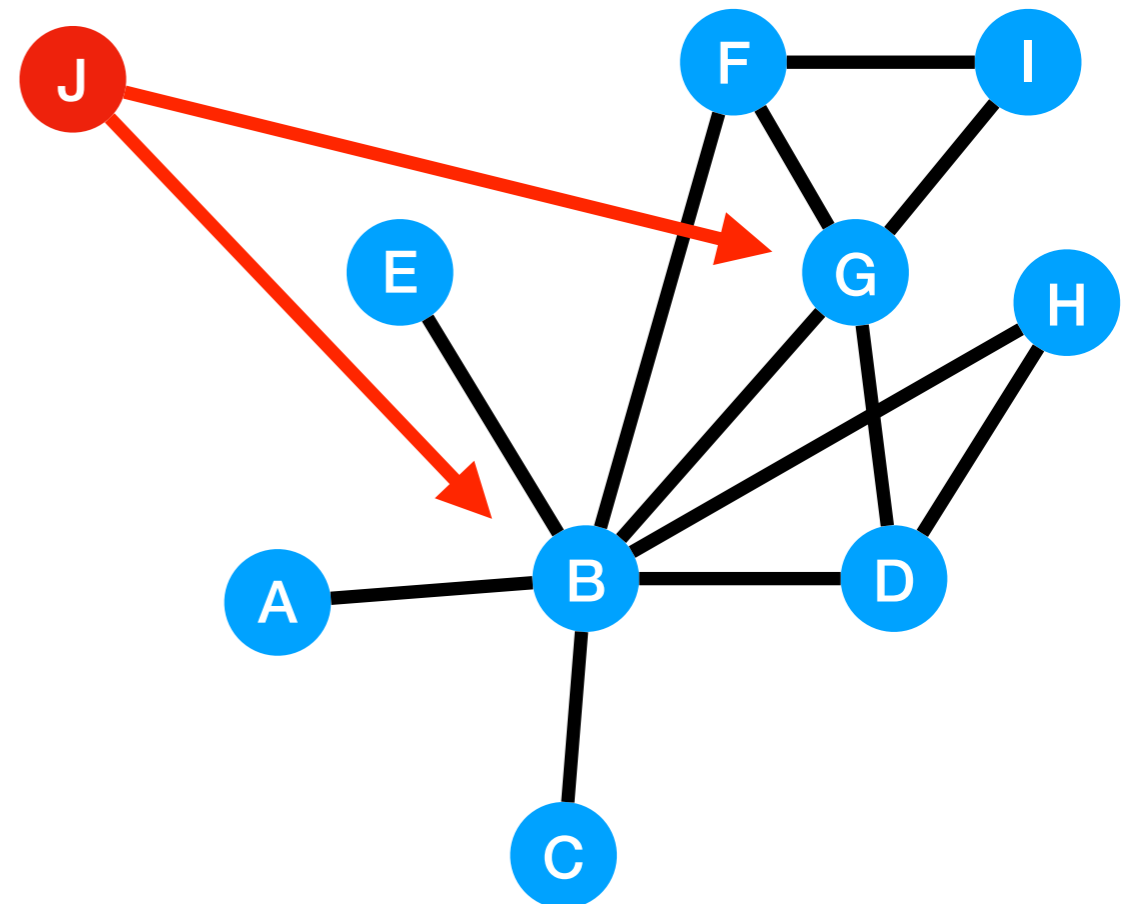
$$\pi_k \sim k^\alpha, \text{ gdzie } \alpha > 1 \implies \text{jeden wierzchołek ma stopień } k \sim O(n)$$

drzewo wygląda jak ‘krzak’



Preferencyjne przyłączanie wierzchołka z m krawędziami

$$\pi_k = \frac{k}{\sum_i k_i} \implies p_k = \frac{4m}{k(k+1)(k+2)}$$



Grafy losowe 1959

Internet pod lupą: 1998-1999

Empiryczne badania innych sieci rzeczywistych

Sieci małego świata + sieci bezskalowe

Sieci rosnące z preferencyjnym przyłączaniem

Związek architektury sieci złożonych z funkcją układu

Complex network science