

Połączeni

W jaki sposób sieci zmieniają nasz sposób patrzenia na świat.

Mikołaj Morzy

- Grafy losowe i model Erdösa-Renyi'ego
 - Sześć stopni separacji
 - Koncentratory i autorytety w sieciach społecznościowych
 - Dołączanie preferencyjne i sieci bezskalowe
 - Błędy, katastrofy i choroby w sieciach
-

**Kto jest największą gwiazdą
Hollywood?**

The Oracle of Bacon

<http://oracleofbacon.org/>



□ Odległość od KB liczona w filmach

Bacon Number	liczba osób
0	1
1	2512
2	263345
3	846332
4	207260
5	15601
6	1444
7	176
8	29

średni BN wynosi 2.981
liczba aktorów: 1336700
średnia liczba filmów: 27

ale...

41% aktorów ma mniej niż 10 filmów

Kevin Bacon nie jest wyjątkowy

□ Analizie poddano IMDB.com

- Kevin Bacon grał z 1800 innymi aktorami w 72 filmach
 - David Carradine grał z 4000 aktorów w 229 filmach
 - Robert Mitchum grał z 2900 aktorami w 135 filmach

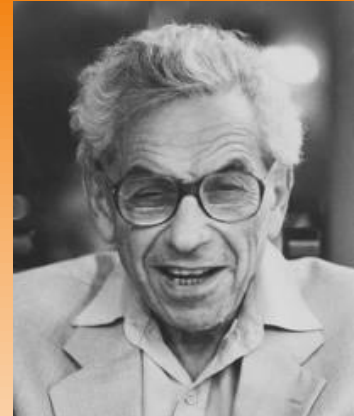
 - odległości między aktorami: R. Steiger (2.53) D. Pleasance (2.54), M. Sheen, Ch. Lee, R. Mitchum, Ch. Heston (2.57)
 - **Kevin Bacon jest na liście 876-ty**

 - najwięcej filmów: M. Blanc (1035), T. Byron (1563), P. North (1616), J.Silvera (968), R. Jeremy (1275)
-

Skąd się biorą koncentratory?

- zarówno model losowy Erdösa-Renyi'ego jak i model małych światów Watts-Strogatza nie pozwalają na istnienie koncentratorów
 - jeśli WWW byłaby siecią losową, prawdopodobieństwo znalezienia strony o 500 linkach przychodzących wynosiłoby ok. 10^{-99}
 - jeśli sieć aktorów byłaby siecią losową, prawdopodobieństwo istnienia Roda Steigera wynosiłoby ok. 10^{-120}
 - we Wszechświecie występuje ok. 10^{80} atomów...
-

Paul Erdős (1913-1996)



- ❑ autor największej liczby artykułów w dziedzinie matematyki (ok. 1500)
 - ❑ całe życie podróżował z jedną walizką
 - *"mój mózg jest otwarty"*
 - ❑ o kawie
 - *"matematyk to maszyna zmieniająca kawę w twierdzenia"*
 - ❑ o amfetaminie (po wygraniu \$500)
 - *"wcześniej patrząc na białą kartkę papieru widziałem setki idei, teraz widzę białą kartkę papieru"*
-

Liczba Erdösa

□ Liczba Erdösa to odległość "współpracownicza" liczona współautorstwem prac naukowych

Liczba Erdosa	Liczba osób
1	1
2	504
3	6593
4	33605
5	83642
6	87760
7	40014
8	11591
9	3146
10	819
11	244
12	68
13	23

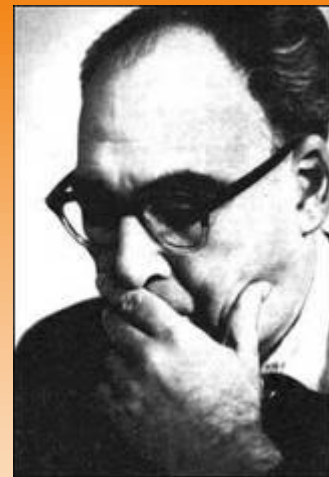
MathSciNet zawiera 1.9 mln publikacji napisanych przez 400 000 osób

50 000 autorów ma nieskończoną Liczbę Erdösa

84 000 nigdy nie opublikowało wspólnej pracy.

źródło: <http://www.oakland.edu/enp/trivia/>

Alfred Renyi (1920-1971)



- ❑ wielki węgierski matematyk, twórca Instytutu Badań Matematycznych Węgierskiej Akademii Nauk
 - ❑ współautor 32 publikacji z Paulem Erdősem
 - ❑ o matematyce
 - *"gdy jestem nieszczęśliwy, zajmuję się matematyką aby stać się szczęśliwy, gdy jestem szczęśliwy, zajmuję się matematyką by pozostać szczęśliwy"*
-

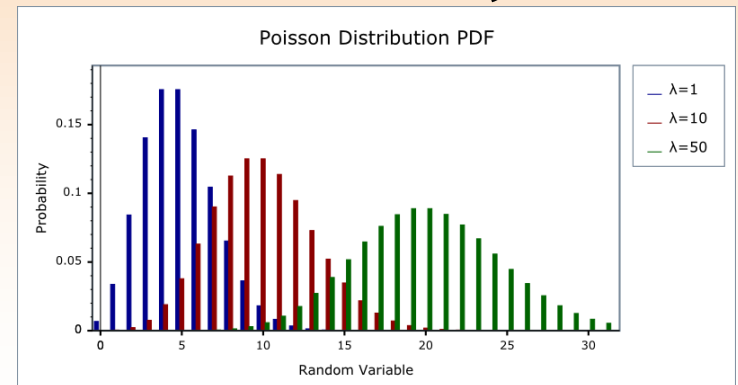
Model grafów losowych

- uniwersalny model powstawania losowych grafów
 - mimo losowego powstawania krawędzi każdy węzeł będzie miał w przybliżeniu ten sam stopień (rozkład stopni wierzchołków to rozkład Poissona)

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

- konsekwencje

- każdy ma mniej więcej tyle samo znajomych
- każda firma współpracuje z tą samą liczbą partnerów
- każdy serwis ma (mniej więcej) tyle samo odsłon



źródło: www.boost.org

Zasada generacji grafu losowego

□ Model I

- **G(n,M)**: graf jest wybierany ze zbioru wszystkich grafów o n wierzchołkach i M krawędziach zgodnie z rozkładem jednostajnym

□ Model II

- **G(n,p)**: w grafie o n wierzchołkach każda krawędź między parą wierzchołków jest tworzona z tym samym prawdopodobieństwem p niezależnie od innych krawędzi

$$p^M (1-p)^{\binom{n}{2}-M} \leftarrow \text{prawdopodobieństwo wygenerowania grafu o } n \text{ wierzchołkach i } M \text{ krawędziach}$$

Narodziny spójnego komponentu

□ w sieciach losowych zachodzi zjawisko perkolacji

$n \cdot p < 1$: brak komponentów większych niż $O(\log n)$

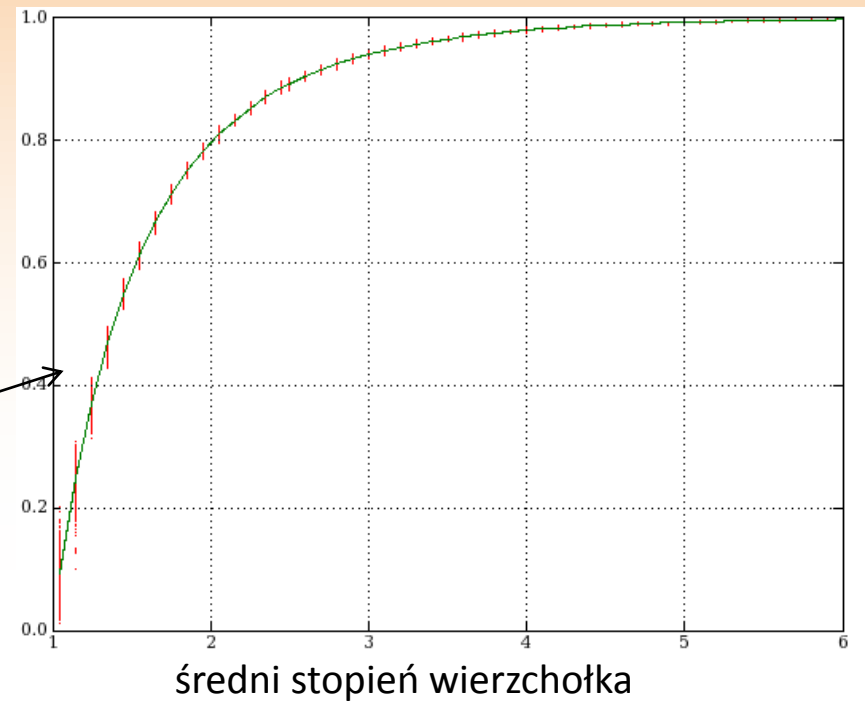
$n \cdot p = 1$: komponent o rozmiarze rzędu $n^{2/3}$

$n \cdot p = \text{const} > 1$: jeden wielki komponent

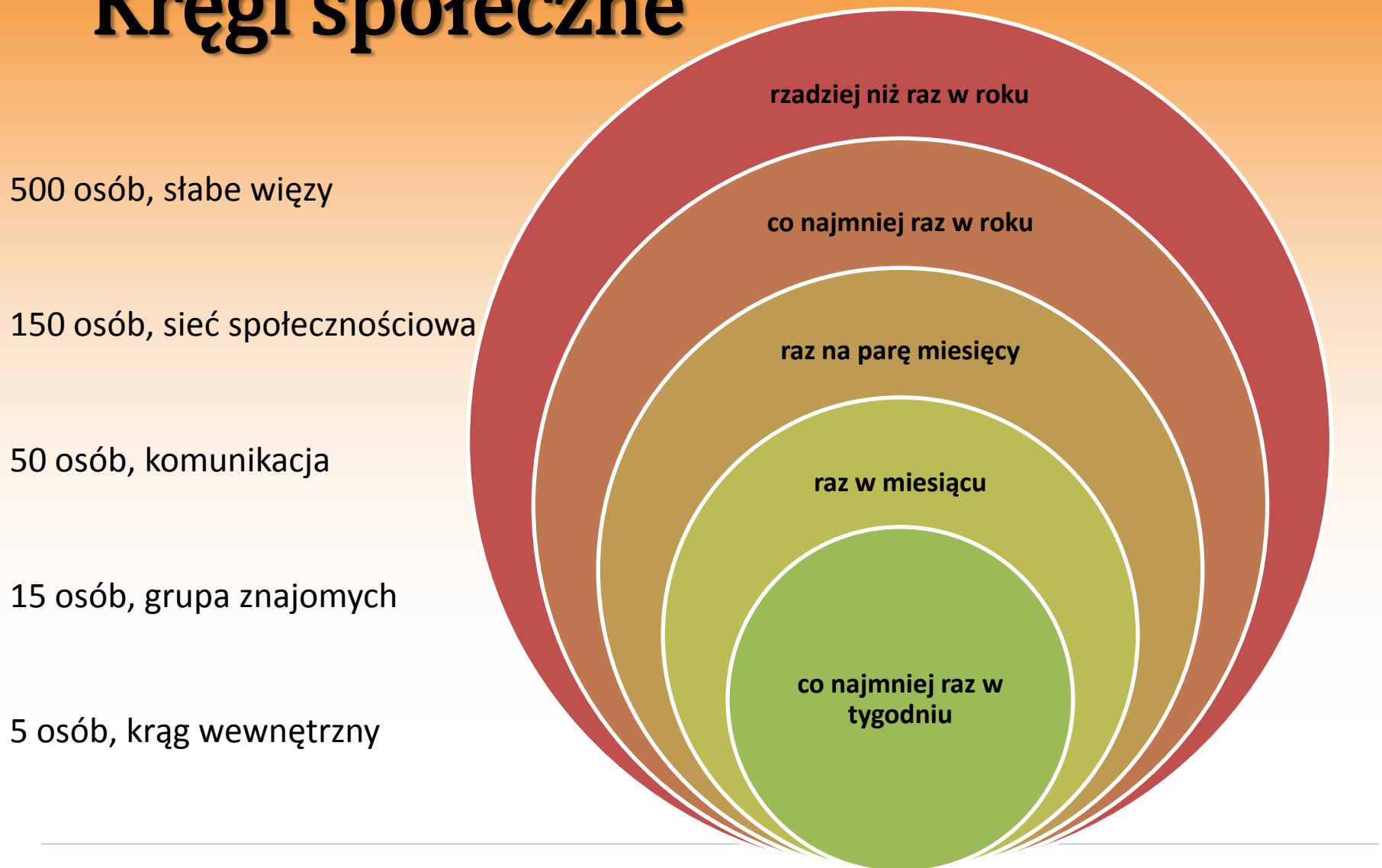
$p < \frac{(1 - \varepsilon) \ln n}{n}$ graf niespójny

$p > \frac{(1 + \varepsilon) \ln n}{n}$ graf spójny

duże odchylenia
w trakcie formowania
się spójnego komponentu



Kręgi społeczne



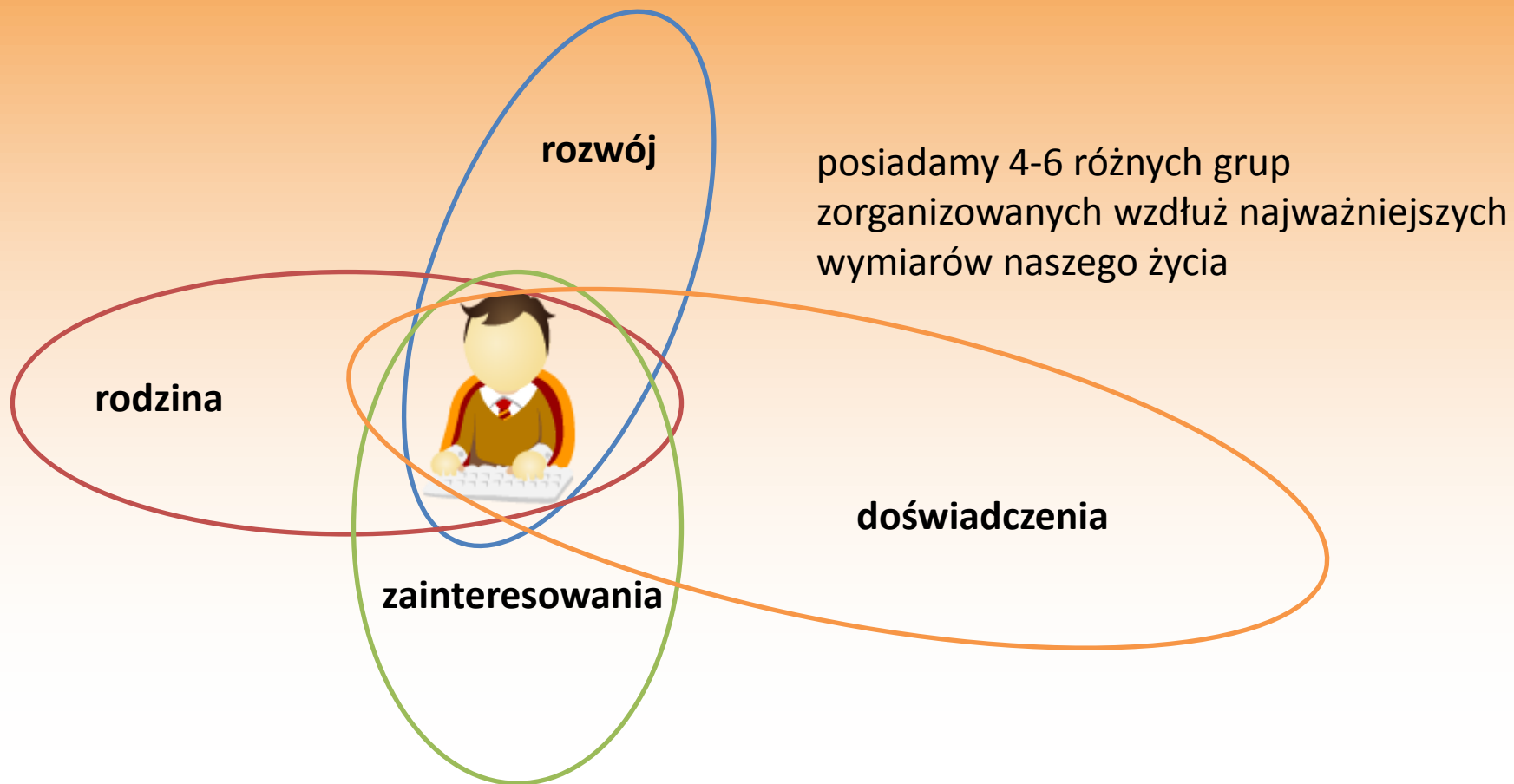
Liczba Dunbara

- Podstawowa sieć społecznościowa składa się ze 150 osób
 - podział wsi w neolicie
 - kohorty rzymskich legionów
 - liczba administratorów artykułu w Wikipedii
 - spójność gier online
 - liczba dni zwolnienia chorobowego w firmie



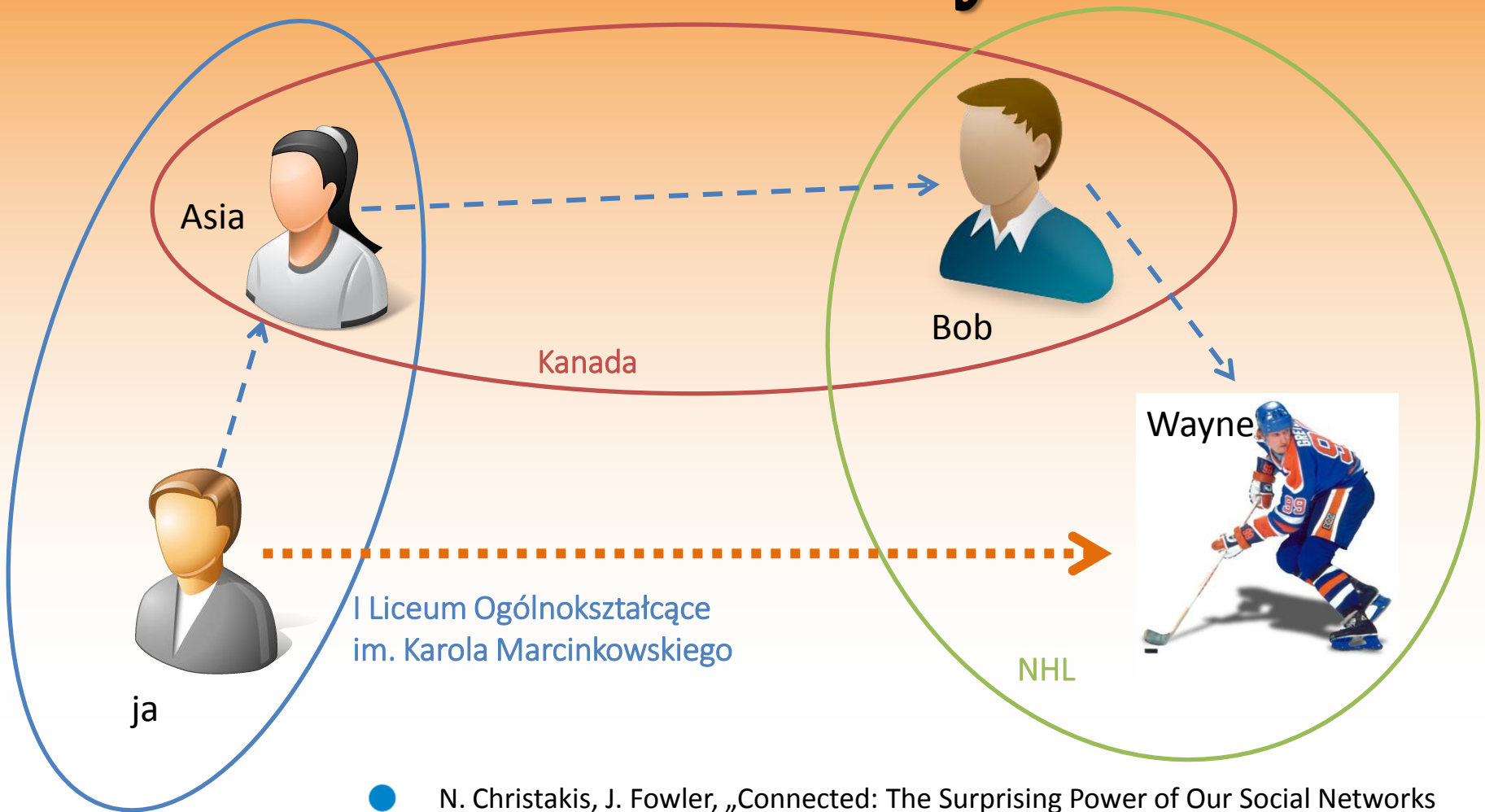
Ch. Allen, „*The Dunbar number as a limit to group sizes*”, blog Life With Alacrity

Jak są zorganizowane kręgi społeczne?



A-L. Barabasi, „Linked: How Everything is Connected to Everything Else and What it Means”, Plume, 2003

Unikalne zbiory ludzi



N. Christakis, J. Fowler, „Connected: The Surprising Power of Our Social Networks and How They Shape Our Lives”, Little, Brown, 2009

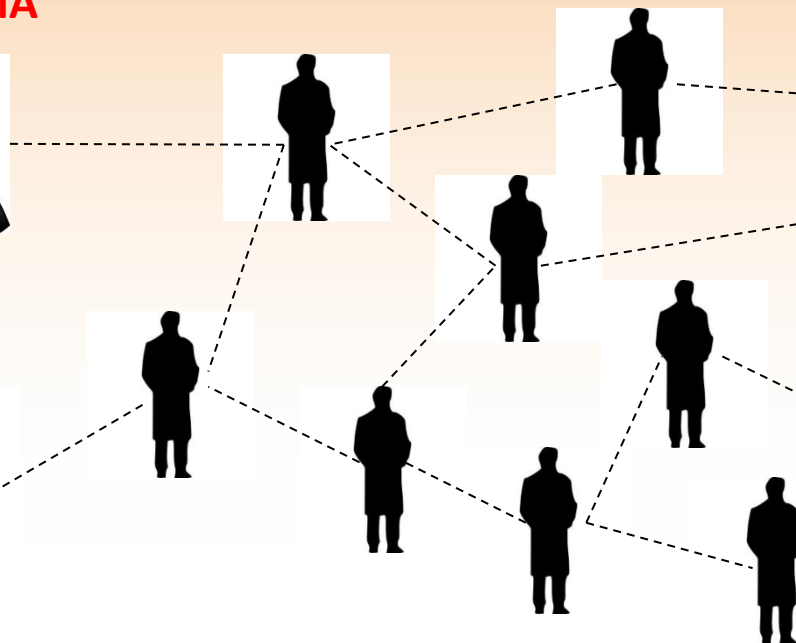
Mit sześciu stopni oddalenia

- ❑ 1929, Frigyes Karinthy, "Láncszemek"
- ❑ 1967, Stanley Milgram, odległość między mieszkańcami USA

Sharon, MA



Boston



Reguły eksperymentu

1. dodaj swoje imię i nazwisko na dole listy nadawców
 2. odepnij pocztówkę i wyślij ją na Uniwersytet Harvarda
 3. jeśli znasz adresata, prześlij list bezpośrednio
 4. jeśli nie znasz adresata osobiście, nie próbuj się kontaktować, wyślij przesyłkę do znajomej osoby która, Twoim zdaniem, może znać adresata
- pierwszy list doszedł po trzech dniach i dwóch przystankach
 - 42 ze 160 listów dotarło do adresatów
 - mediana pośredników wynosiła 5.5
 - Milgram nigdy nie użył terminu "*six degrees of separation*"

Everybody on this planet is separated by only six other people. Six degrees of separation. Between us and everybody else on this planet. The president of the USA. A gondolier in Venice... It's not just the big names. It's anyone. A native in the rain forest. A Tierra del Fueagan. An Eskimo. I am bound to everyone on this planet by a trail of six people. It's a profound thought. How every person is the new door opening up into other worlds.

John Guare, "*Six degrees of separation*"

Rozmiary sieci WWW

- ❑ czy wynik eksperymentu Milgrama można odnieść do sieci WWW?
 - w 1999 r. Albert-László Barabási i Réka Albert postanowili sprawdzić wielkość sieci
 - średnica wyniosła 19 linków
 - dziś średnica sieci wynosi 21.7 a aktualna liczba stron to ok. 50 mld

$$d = 0.35 + 2 * \log n \quad \text{w rzeczywistej sieci WWW}$$

$$d = \log n / \log k \quad \text{w sieci losowej}$$

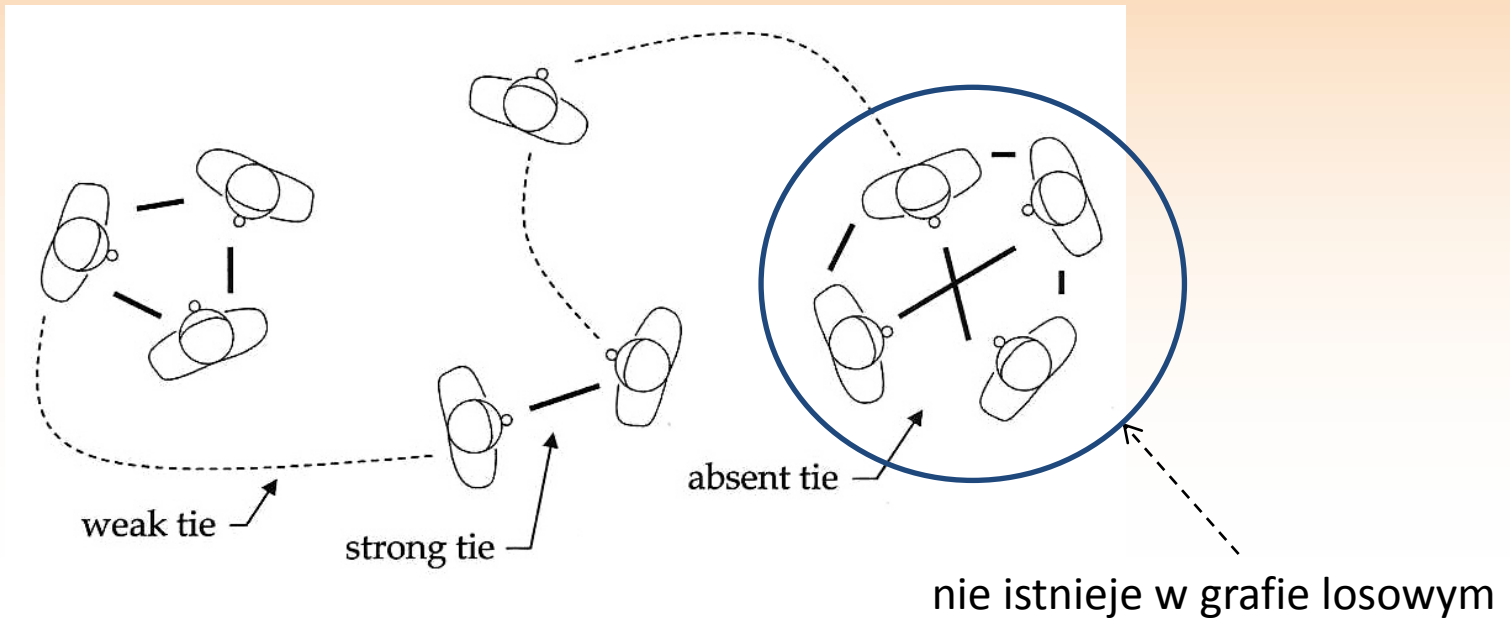


Réka Albert, Hawoong Jeong, and Barabási, Albert-László (1999). "The Diameter of the WWW". *Nature* **401** (6749): 130–131. [arXiv:cond-mat/9907038](https://arxiv.org/abs/cond-mat/9907038). [Bibcode 1999Natur.401..130A](https://doi.org/10.1038/43601). [doi:10.1038/43601](https://doi.org/10.1038/43601).

Jak znaleźć pracę?



- Podstawowa koncepcja
 - silne więzy przyjaciół
 - słabe więzy znajomych



Mark S. Granovetter, "The Strength of Weak Ties"
American Journal of Sociology **78** (6) (May, 1973), pp. 1360-1380

Co wspólnego mają świerszcze i teatr?

- zjawisko synchronizacji w sieci
 - klaskanie w teatrze
 - rozbłyski azjatyckich świetlików
 - cykanie świerszczy
 - *C.elegans* i jego system nerwowy

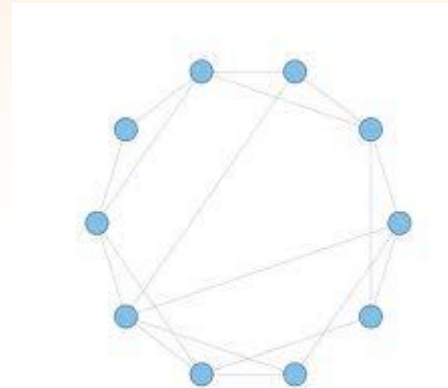
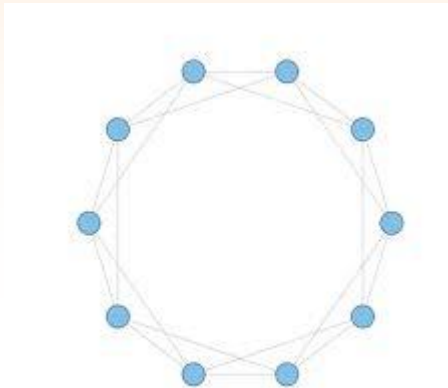
- współczynnik grupowania (ang. *clustering coefficient*) jest w sieciach rzeczywistych 10 000 razy większy niż w sieciach losowych.



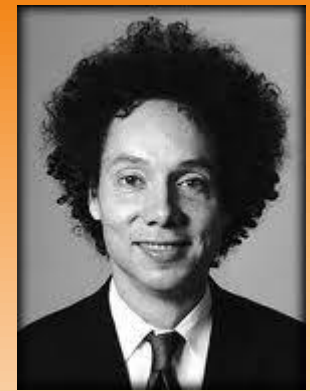
Watts & Strogatz

Model Watts-Strogatz

- ❑ ludzie żyją na okręgu, każdy zna swoich bezpośrednich 4 sąsiadów
 - współczynnik grupowania $cc = 0.5$ (w sieci losowej współczynnik zależy od liczby węzłów)
 - losowo osoby są łączone krawędzią (z małym prawdopodobieństwem)
 - dodane krawędzie nie wpływają na współczynnik grupowania
 - dodane krawędzie powodują powstanie efektu "małych światów"
 - ... ale dane nie zgadzają się z tym modelem



The Tipping Point



- ❑ Malcolm Gladwell, autor *New Yorkera*, publikuje książkę zawierającą prosty test stopnia socjalizacji
 - 248 nazwisk losowo wybranych z książki telefonicznej
 - 1 punkt za każde rozpoznane nazwisko

grupa	średnia	rozrzut
studenci	21	2 – 95
profesorowie	39	9 – 118



Reguła 80/20



- Vilfredo Pareto (1848-1923), włoski inżynier, socjolog, naukowiec i filozof
 - zauważył, że 80% groszku było produkowane przez 20% krzaków
 - 80% ziemi we Włoszech należało do 20% rodzin włoskich
 - 80% zysków jest generowanych przez 20% pracowników
 - 80% reklamacji jest zgłaszanych przez 20% klientów
 - 80% decyzji jest podejmowanych w trakcie 20% spotkania

 - 80% linków WWW kieruje do ok. 15% stron
 - 80% cytowań dotyczy 38% naukowców
 - 80% odnośników w Hollywood wskazuje na 30% aktorów
-

Konsekwencje reguły 80/20

- ❑ reguła 80/20 znamionuje procesy rządzone rozkładem potęgowym, w którym element losowy gra niewielką rolę
 - ❑ pojawienie się rozkładu potęgowego najczęściej oznacza przejście od chaosu do stanu uporządkowanego
 - przejście fazowe (woda-lód, ferromagnetyki)
 - w pobliżu punktu krytycznego wiele kluczowych wartości zaczyna być charakteryzowanych rozkładem potęgowym
 - 1971, K.Wilson, renormalizacja

 - ale jaki jest mechanizm pojawiania się praw potęgowych?
-

Efekt św. Mateusza

Każdemu bowiem, kto ma, będzie dodane, tak że nadmiar mieć będzie. Temu zaś, kto nie ma, zabiorą nawet to, co ma.

❑ Model A

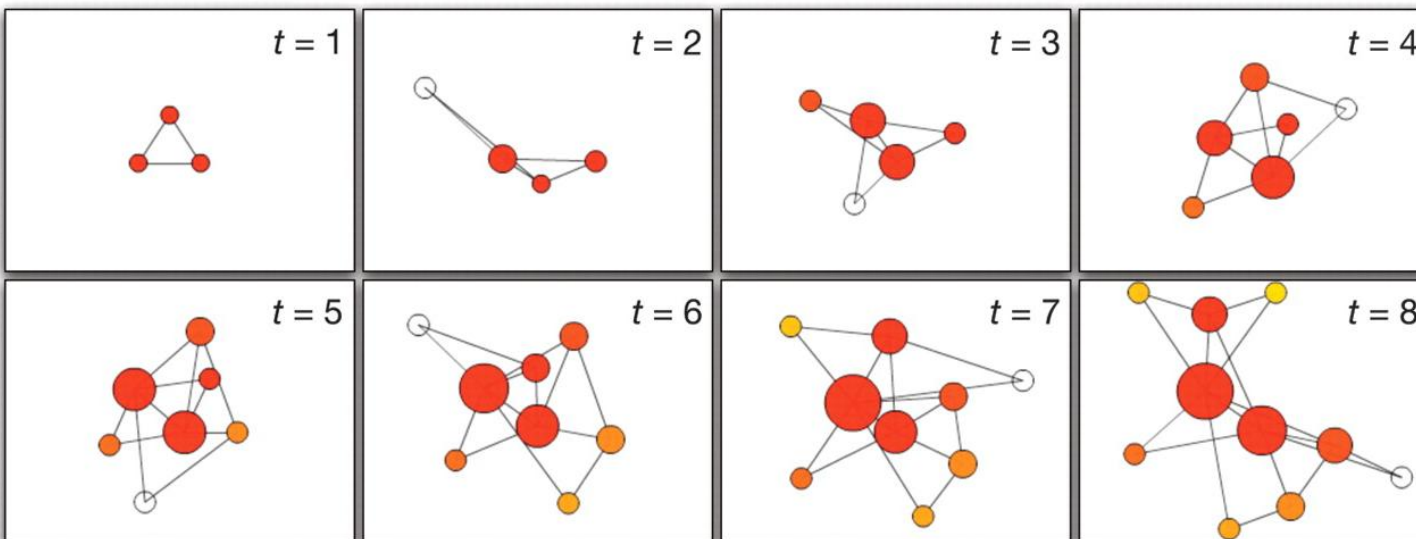
- dodawaj nowe węzły, losowo dołączając je do węzłów, które już istnieją w sieci
- wykładniczy rozkład stopni, sam **wzrost** sieci nie tłumaczy potęgowego charakteru sieci

❑ Model B

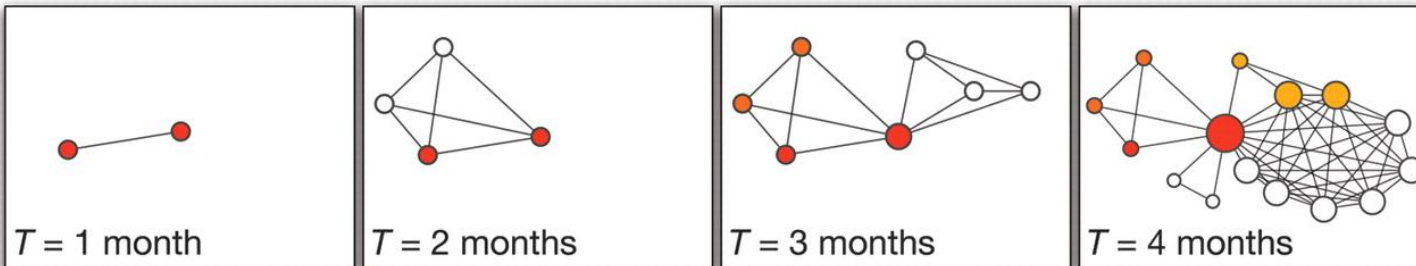
- dodawaj nowe węzły, wybierając węzeł docelowy proporcjonalnie do stopnia węzła
 - potęgowy rozkład stopni, **wzrost i preferencyjne dołączanie** tłumaczą powstawanie potęgowego charakteru sieci
-

Narodziny sieci bezskalowej

Scale-Free Model



Scientific Collaboration Network



źródło: <http://www.sciencemag.org/content/325/5939/412/F1.expansion.html>

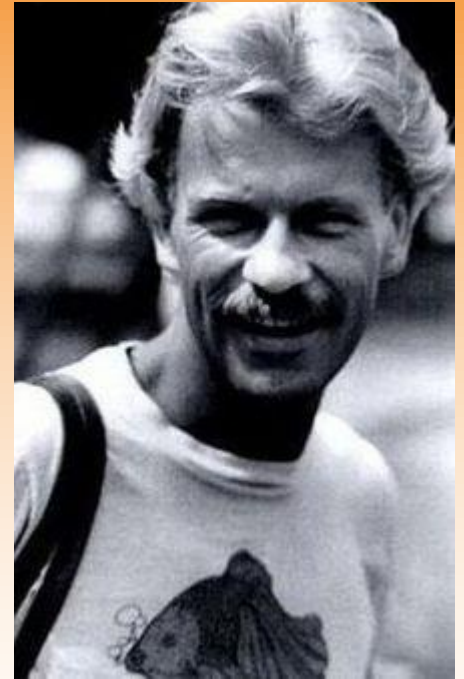
Great Northeast Blackout

- ❑ 9 lutego 1965 roku jedna linia przesyłowa o mocy 230 kV położona w okolicach Ontario pękła
- ❑ w konsekwencji na 24h utracono prąd w mieście Nowy Jork i 7 stanach
 - 800 tys. osób uwięzionych w metrze
 - 5 000 żołnierzy i policjantów
 - 30 mln osób odczuło awarię

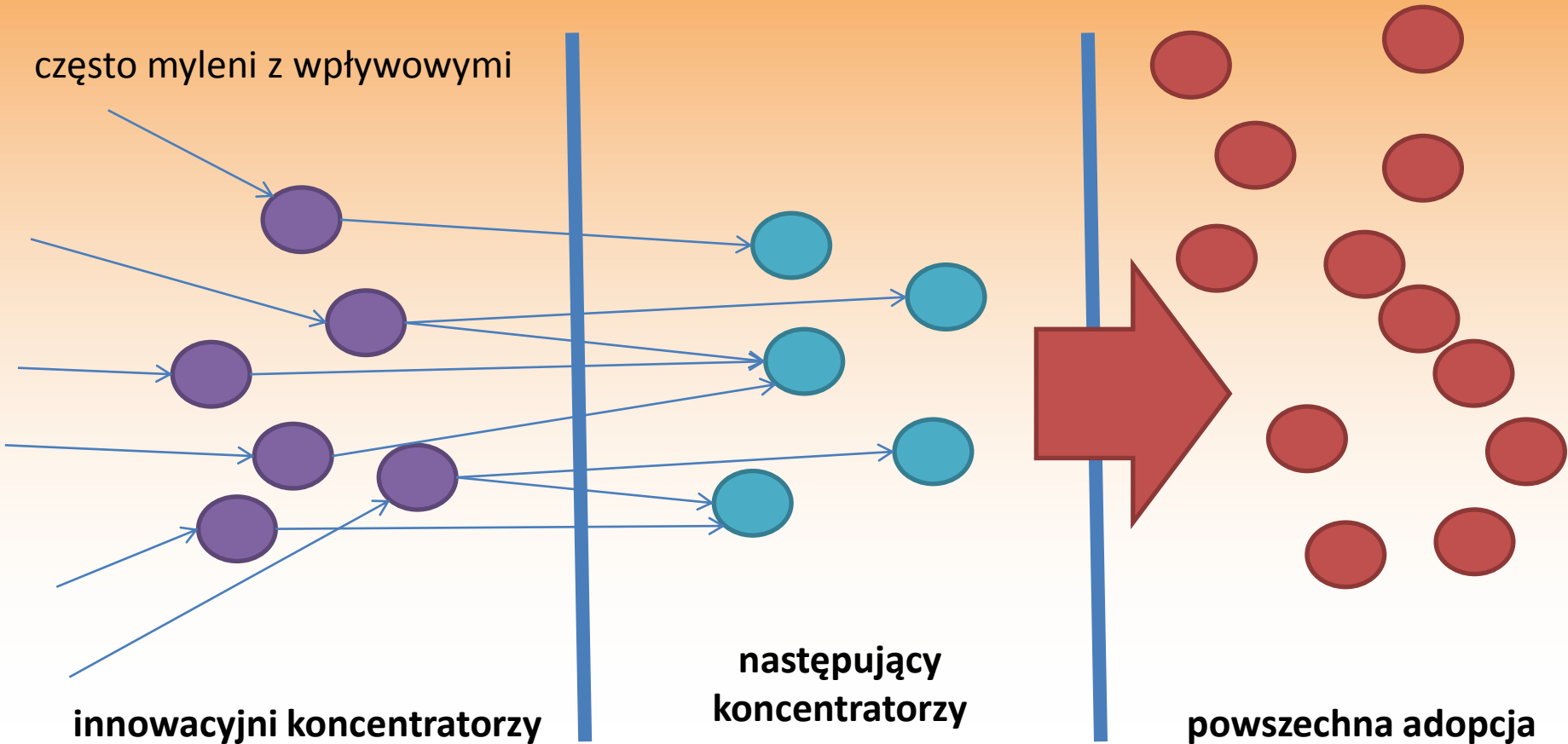


Pacjent Zero

- ❑ Gaetan Dugas (1953-84)
- ❑ steward w samolotach Air Canada
- ❑ 40 z 248 osób zdiagnozowanych z AIDS do kwietnia 1982 roku uprawiało z nim seks
 - około 250 partnerów rocznie
 - około 2500 partnerów w ciągu życia
 - bohater książki "*And the band played on*"



Proces adopcji



D. Watts, „Everything is Obvious: Once You Know the Answer”
Crown Business, 2011

Co na nas wpływa?

- wszystko co znajduje się w naszym sąsiedztwie
 - jedzenie w dwie osoby zwiększa spożycie o 35%
 - jedzenie w cztery osoby zwiększa spożycie o 75%
 - przyjaciele wpływają na palenie, naukę, zakupy, nastrój...
 - wszystko co zdarzyło się wcześniej
 - D. Watts, eksperyment z wyborem muzyki
 - wszystko co robią osoby z naszej grupy
 - identyfikacja członka grupy zajmuje ok. 170 ms
-

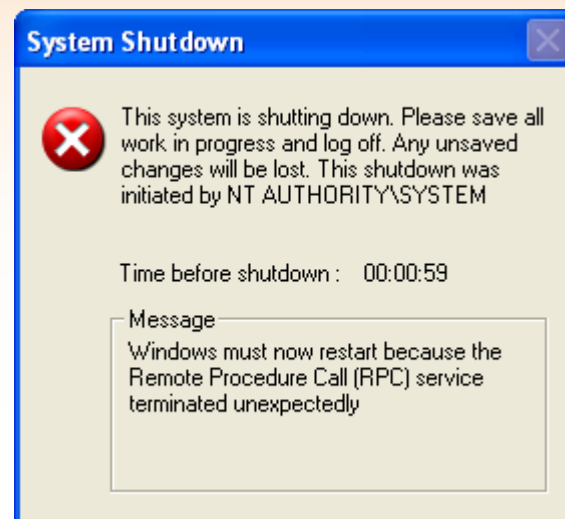
Niestety, nie tylko idee się roznoszą

□ Poczta elektroniczna

- list od Clarie Swire do Bradleya Chaita
- 7 mln czytelników w przeciągu tygodnia

□ Wirusy komputerowe

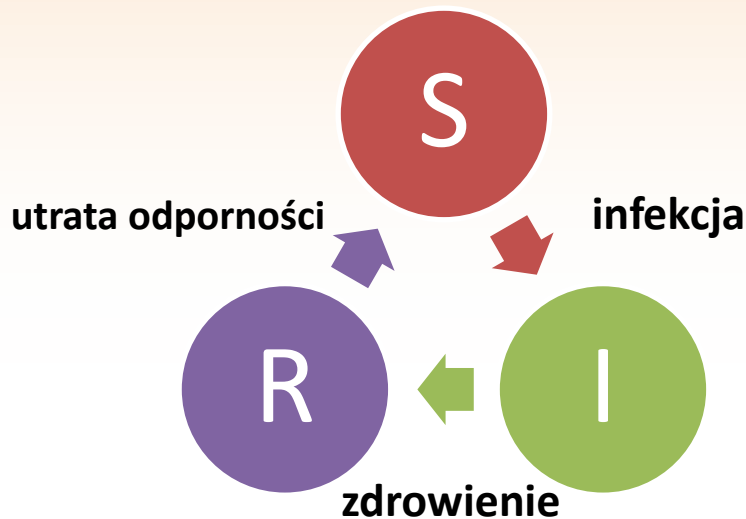
- Marzec 1999, wirus Melissa
- 100 000 komputerów w 300 organizacjach przez weekend



Tradycyjny model infekcji

□ Fazy infekcji

- *susceptible*: podatny, może być zainfekowany
- *infectiuous*: zarażony, aktywnie infekuje innych
- *removed*: usunięty, wyleczony lub martwy



zakłada się, że częstotliwość kontaktów między członkami grup S, R, I jest losowa

W. Kermack, A.G. McKendrick

Logistyczny model infekcji

- faza powolnego wzrostu
- faza eksplozji
- faza wypalenia



zarażonych



$$\text{logit}(p) = \log\left(\frac{p}{1-p}\right) = \log(p) - \log(1-p)$$

$$\text{logit}^{-1}(p) = \frac{1}{1+e^{-p}} = \frac{e^p}{e^p+1}$$

tempo reprodukcji: jak wiele infekcji jest tworzonych przez jednego zarażonego osobnika

Model SNA infekcji

□ Choroba bierze pod uwagę topologię sieci

- zarażenia na „frontach choroby” na granicach grup
- losowe linki umożliwiają rozprzestrzenianie choroby
- wyjaśnia czarna śmierć, HIV, pryszczycę, ...

