



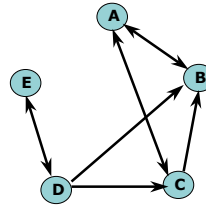
## Sieci społeczne

Charakterystyka, uwarunkowania i konsekwencje struktur relacji społecznych na przykładzie komunikacji internetowej

dr Dominik Batorski  
Instytut Socjologii, UW

Praca doktorska napisana pod kierunkiem  
Prof. dr hab. Andrzeja Nowaka

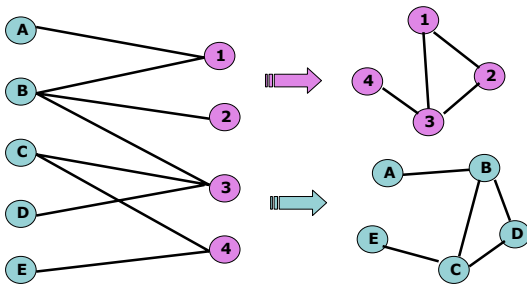
## Sieci: grafy i macierze



	A	B	C	D	E
A	-	1	1	0	0
B	1	-	0	0	0
C	1	1	-	0	0
D	0	1	1	-	1
E	0	0	0	1	-

2

## Sieci afiliacji



3

## Analiza sieci społecznych

- ◆ Charakterystyki pozycji jednostek w sieci
  - Centralność
  - Prestiż
- ◆ Grupy w obrębie sieci
  - Kliki
  - Pozycje strukturalne
- ◆ Globalne własności sieci

4

## Najważniejsze pytania

- ◆ Struktura sieci personalnych
  - grupy czy sieci?
- ◆ Własności struktury dużych sieci
  - Komponenty globalne
  - Problem „małego świata”
  - Rozkład liczby relacji
  - Mixing patterns

5

## Komunikatory internetowe

- ◆ Komunikatory to wygodny sposób na komunikowanie się przez Internet w czasie rzeczywistym.
- ◆ Umożliwiają sprawdzenie obecności innych osób w Sieci.
- ◆ Krótkie wiadomości tekstowe.
- ◆ Inne funkcje:
  - Wysyłanie SMS,
  - Przesyłanie plików.
  - Rozmowy głosowe.
  - Tryb konferencyjny.



6

## Gadu-Gadu

- ◆ Gadu-gadu (GG) to najpopularniejszy komunikator w Polsce.
- ◆ Obecnie ma około 3 milionów użytkowników.
- ◆ Komunikatory są używane:
  - Przede wszystkim do kontaktu ze znajomymi, przyjaciółmi i rodziną;
  - czasem również do podtrzymywania kontaktu z osobami poznanymi w Sieci.



7

## Zgromadzone dane

- ◆ Informacje o relacjach z list kontaktów:
  - Dane ze stycznia 2004;
  - Listy 3,354,457 aktywnych użytkowników;
  - Informacja o tym kto ma kogo w swojej skrzynce.
  - Zawierające, ponad 75 milionów relacji;
  - W sumie 5,390,453 zarejestrowanych użytkowników. I prawie 3 miliony posiadających odwzajemnione relacje.
- ◆ Dane indywidualne:
  - Płeć, wiek, miejsce zamieszkania.
- ◆ Dane o komunikacji.
  - Dane z czterech tygodni maja 2004;
  - Informacja o tym kto w danym tygodniu do kogo wysłał wiadomości.

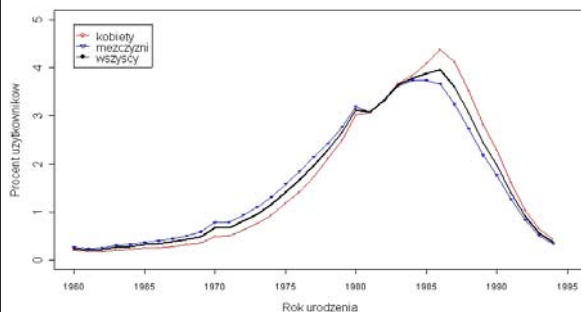
8

## Rodzaje analizowanych sieci

1. Sieć relacji z list kontaktów;
2. Sieci komunikacji krótkookresowej (poszczególne tygodnie);
3. Sieci komunikacji długookresowej (miesiąc);
4. Sieci intensywnej komunikacji (przynajmniej raz w tygodniu przez miesiąc);
5. Sieci relacji długotrwałych (istniejących minimum 4 miesiące);
6. Sieci mocnych relacji (odwzajemnionych, długotrwałych i intensywnych);

9

## Płeć i wiek użytkowników



10

## Liczba relacji

Rodzaj relacji		Średnia	Max
<b>odwzajemnione</b>		<b>14.2</b>	<b>430</b>
przychodzące		22.1	9466
wychodzące		25.0	1000
powiązania		32.9	9481

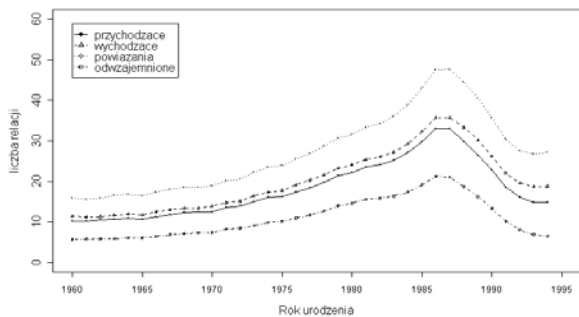
11

## Relacje a płeć

- ◆ Informacje o płci podaje 87% użytkowników
  - 37% stanowią kobiety
  - 49% mężczyźni
- ◆ Mężczyźni mają więcej relacji niż kobiety.
- ◆ Relacje pomiędzy osobami różnej płci są bardziej prawdopodobne.
- ◆ Jednak odwzajemnienie relacji jest większe w obrębie płci.

12

## Relacje a wiek



13

## Badania internetu

- ◆ Internet jest wykorzystywany przede wszystkim w celach związanych z aktywnościami życia codziennego.
- ◆ Komunikacja w Internecie odbywa się głównie pomiędzy osobami, które się znają i utrzymują ze sobą relacje również poza Internetem.
- ◆ Wiele z prawidłowości dotyczących interakcji obserwowanych w Internecie jest bardzo podobnych do prawidłowości obserwowanych poza Internetem.
- ◆ Błędem jest traktowanie Internetu jako zupełnie nowej i niezależnej rzeczywistości społecznej.

14

## Grupy vs. sieci

- ◆ Formalnie grupa jest jednym z możliwych rodzajów sieci. Łatwiej jest jednak porównywać metaforę grup z metaforą sieci.
- ◆ Każda osoba tworzy raczej własną sieć personalną niż należy do jednej grupy z osobami, z którymi jest w relacjach.
- ◆ Choć ludzie bardzo często widzą świat w terminach grup, to jednak funkcjonują w relacjach, których struktury mają znacznie bardziej sieciowy charakter.

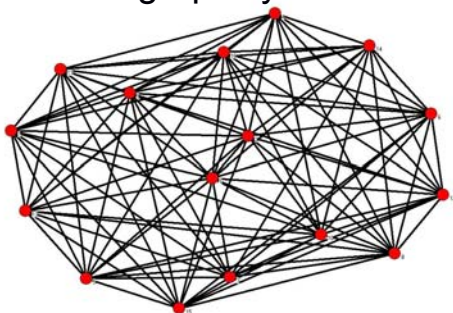
15

## Grupy vs. sieci

- |                                |                                   |
|--------------------------------|-----------------------------------|
| ◆ Grupy                        | ◆ Sieci                           |
| • duża gęstość                 | • niewielka gęstość               |
| • wyraźne granice              | • brak wyraźnych granic           |
| • zorganizowane hierarchicznie | • fragmentaryczność               |
| • trwałe                       | • nietrwałe                       |
| • homogeniczne                 | • zróżnicowane                    |
|                                | • większa rozpiętość przestrzenna |

16

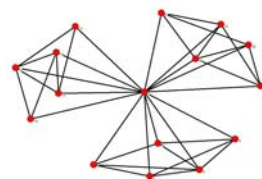
## Struktury o charakterze grupowym



17

## Struktura sieciowa (1)

- ◆ Jednostka jako członek wielu niezależnych grup.



18

## Struktura sieciowa (2)

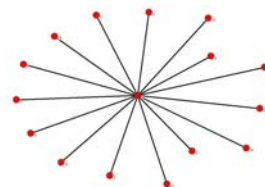
- ◆ Struktura sieciowa.
  - Niska gęstość



19

## Struktura sieciowa (3)

- ◆ Gwiazda socjometryczna – brak relacji pomiędzy znajomymi ego.



20

## Sieci personalne

- ◆ Sieciowy indywidualizm (Wellman)
- ◆ Miary:
  - Lokalna gęstość
  - Komponenty lokalne
  - Wspólni znajomi (pokrywanie się sieci personalnych)

21

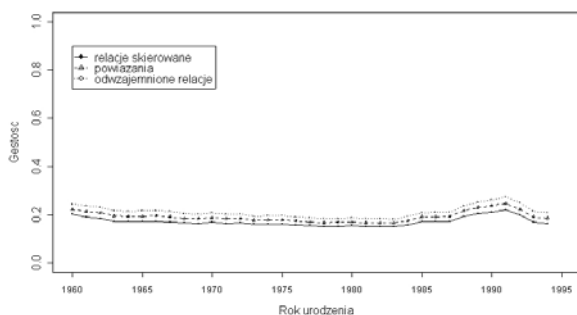
## Lokalna gęstość

- ◆ Lokalna gęstość, relacje między znajomymi

	Średnia
Relacje skierowane	<b>0.206</b>
Powiązania	<b>0.225</b>
Odwzajemnione	<b>0.186</b>

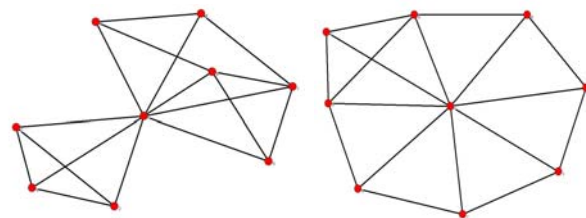
22

## Gęstość w zależności od wieku



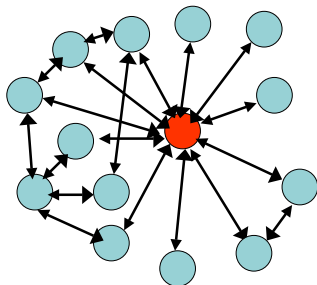
23

## Znaczenie lokalnych komponentów



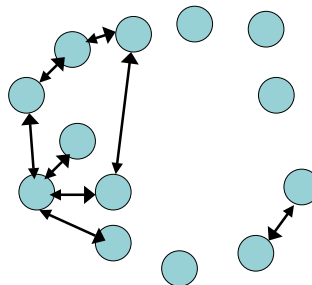
24

## Sieci personalne



25

## Sieci personalne



26

## Komponenty lokalne

	Max.	Średnia
Liczba komponentów	343	<b>5.595</b>
Wielkość największego komp.	365	<b>8.514</b>
Udział największego komp.	1	<b>0.621</b>
Średnia liczebność pozostałych komp.	28	<b>1.219</b>

27

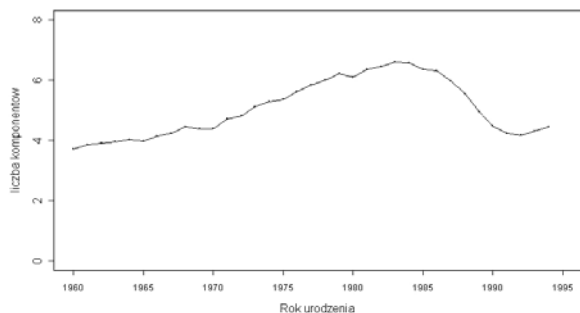
## Duże sieci personalne

- ◆ Użytkownicy posiadający 10 i więcej odwzajemnionych relacji.
  - 35% aktywnych użytkowników.
- ◆ posiadają przeciętnie 27 odwzajemnionych relacji, oraz:
 

• Lokalna gęstość	0.13
• Liczba komponentów	9.5
• Udział największego komponentu	56%
• Przeciętny rozmiar pozostałych komp.	1.34

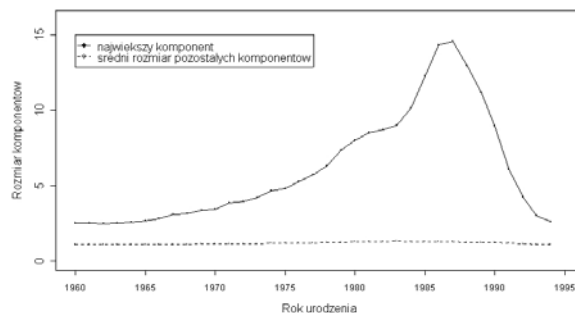
28

## Liczba komponentów i wiek



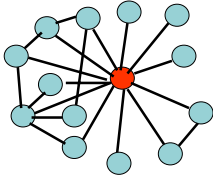
29

## Rozmiar komponentu



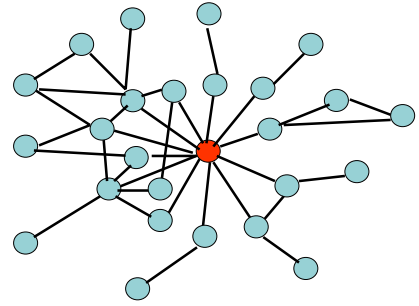
30

## Sieci personalne (promień 1)



31

## Sieci personalne (promień 2)



32

## Znajomi znajomych

- ◆ Sieci personalne uwzględniające znajomych znajomych.
- ◆ Średnia liczebność takiej sieci personalnej to: 398 użytkowników
  - Lokalna gęstość 0,07
  - Liczba komponentów lokalnych 2,5
  - Wielkość największego komponentu 370
  - Udział największego komponentu 87,3%
  - Wielkość pozostałych komponentów 17,9

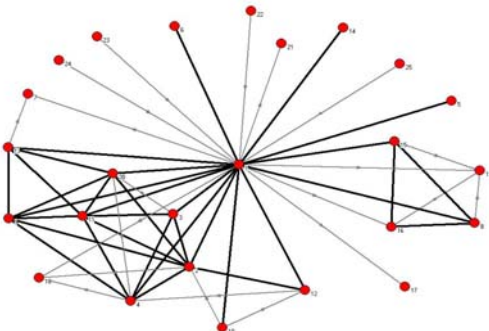
33

## Wspólni znajomi (pokrywanie się sieci personalnych)

- ◆ Średni procent wspólnych znajomych: 4.7%

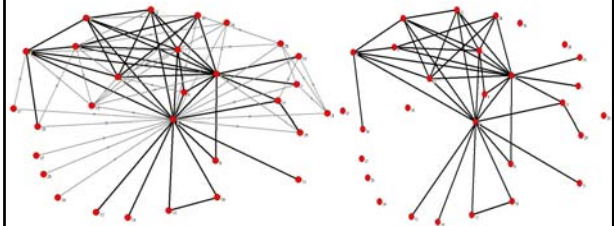
34

## Typowa sieć personalna



35

## Typowa sieć personalna 2



36

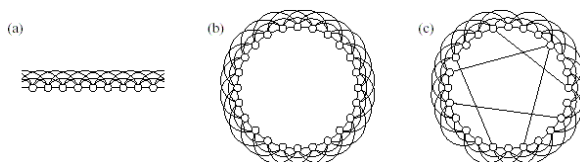
## Problem „małego świata”

- ◆ Eksperymenty Milgrama (1967, 1969)
  - Listy przekazywane od osoby do osoby docierały do docelowej osoby zaledwie w kilku krokach.
- ◆ Wyjaśnienie zjawiska (Watts i Strogatz, 1998)



37

## Model małych światów (Watts i Strogatz, 1998)



- (a) Jednowymiarowa sieć, w której każdy z węzłów połączony jest z sześcioma najbliższymi sąsiadami.
- (b) Ten sam model na okręgu po połączeniu obu końców w celu uniknięcia problemu krawędzi.
- (c) Model Watts'a i Strogatz'a powstały w wyniku przepisania niewielkiej części krawędzi (wybranych losowo z małym prawdopodobieństwem) do losowo wybranych węzłów.

38

## Komponenty

- ◆ (*Component*) Komponent, do którego należy dany węzeł to zbiór węzłów, do którego można dotrzeć poprzez ścieżki relacji pomiędzy węzłami.
- ◆ Ile jest komponentów w sieci?

39

## Komponenty

- ◆ Liczba i rozmiar komponentów – powiązanych podgrafów:
  - Słabe powiązania  
5390453 użytkowników; 1066 komponentów.
- Powiązania odwzajemnione  
2948417 użytkowników; 5706 komponentów.



40

## Słabe połączenia

- ◆ 1066 komponentów

Rozmiar komp.	Liczba komp.
<b>5388245</b>	1
7	1
6	1
4	9
3	51
2	1003



41

## Połączenia odwzajemnione

- ◆ 5706 komponentów
- ◆ **99.57%** użytkowników w największym komponencie

Rozmiar komp.	Liczba komp.
<b>2935753</b>	1
14	2
12	1
8-11	22
5-7	82
4	121
3	523
2	4954



42

## Próba

- ◆ Próba użytkowników dobrana w sposób losowy.
- ◆ Wielkość próby
  - 14782 użytkowników (0.5% populacji);
  - w tym 14727 z największego komponentu.
- ◆ Odległość użytkownika od każdego innego w największym komponentcie.
- ◆ W sumie ponad 43,2 miliardy przeanalizowanych par.

43

## Pytania

- ◆ **Geodesic path:** A geodesic path is the shortest path through the network from one vertex to another
  - (there may be and often is more than one geodesic path between two vertices).
- ◆ **Diameter:** The diameter of a network is the length (in number of edges) of the longest geodesic path between any two vertices.

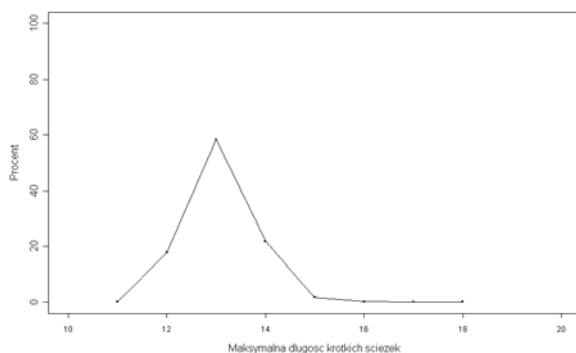
44

## Najkrótsze ścieżki

	Min.	Max.	Średnia
Największa odległość	<b>11</b>	<b>18</b>	<b>13.08</b>
Średnia odległość	4.37	10.61	<b>5.786</b>
Wariancja długości	0.49	0.59	0.539

45

## Średnica sieci



46

## Rozkład liczby relacji

- ◆ Rozkład liczby relacji odbiega znacznie od rozkładu normalnego.
- ◆ Scale-free networks (Barabási, Barabási i Albert)
  - Rozkład potęgowy
    - ◆ power law distribution  $P_k \approx k^{-\alpha}$
  - Rozkład wykładniczy
    - ◆ Exponential distribution  $P_k \approx e^{-k/k}$

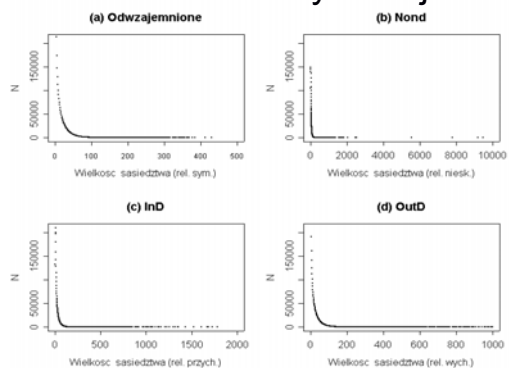
47

## Źródła potęgowego rozkładu liczby relacji

- ◆ Wzrost sieci:
  - Początkowo sieć składa się z niewielkiej liczby węzłów. W każdej jednostce czasu dodawane są nowe węzły. Nowy węzeł jest połączony z m spośród istniejących węzłów.
- ◆ Dołączanie preferencyjne:
  - Prawdopodobieństwo dodania połączenia do istniejącego węzła zależy od liczby relacji posiadanych przez ten węzeł. Im więcej ma on relacji tym większe prawdopodobieństwo, że otrzyma kolejną.

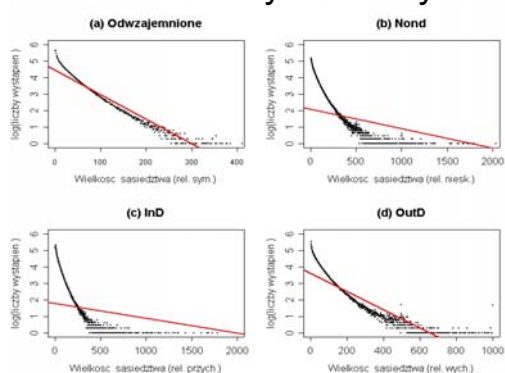
48

## Rozkład liczby relacji



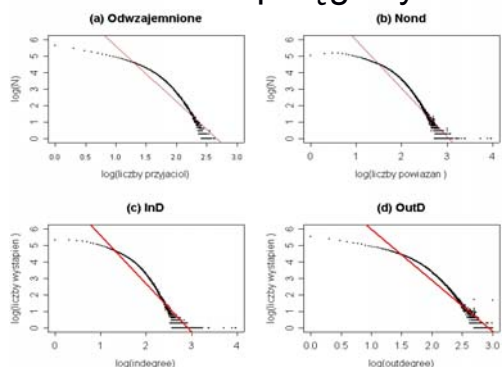
49

## Rozkład wykładniczy



50

## Rozkład potęgowy



51

## Rozkład potęgowy czy wykładniczy?

- ◆ Relacje odwzajemnione
  - $R^2$  dla rozkładu wykładniczego 0.93
  - $R^2$  dla rozkładu potęgowego 0.88
- ◆ Powiązania
  - $R^2$  dla rozkładu wykładniczego 0.27
  - $R^2$  dla rozkładu potęgowego 0.90
- ◆ Rozkład wykładniczy lepiej wyjaśnia rozkład liczby relacji wychodzących, a rozkład potęgowy relacji przychodzących.

52

## Charakter rozkładu liczby relacji: potęgowy czy wykładniczy?

	Rozkład wykładniczy	Rozkład potęgowy
Odwzajemnione relacje	$R^2 = 0.93$	$R^2 = 0.88$
Słabe relacje	$R^2 = 0.27$	$R^2 = 0.90$

53

## Mixing patterns

- ◆ Jakie węzły są ze sobą w relacji?
  - Różne rodzaje (typy) węzłów,
  - prawdopodobieństwo połączenia pomiędzy węzłami zależy od typów węzła
- ◆ Homofilia:
  - Upodobanie do podobieństwa, nakazujące przyjaźnić się z tymi, którzy są do nas podobni wiekiem, zawodem, miejscem zamieszkania, wyznaniem, poglądami, wartościami, światopoglądem.
- ◆ Ten rodzaj selektywnych powiązań jest też nazywany assortative mixing

54

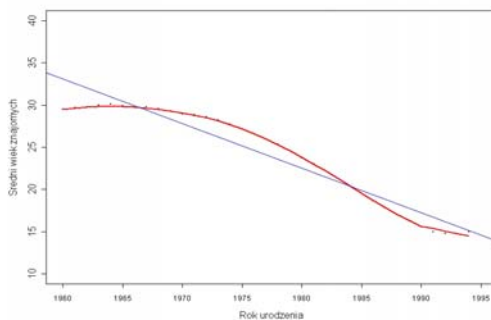
## Homofilia

### ◆ Homofilia:

Upodobanie do podobieństwa, nakazujące przyjaźnić się z tymi, którzy są do nas podobni wiekiem, zawodem, miejscem zamieszkania, wyznaniem, poglądami, wartościami, światopoglądem.

55

## Zależność od wieku



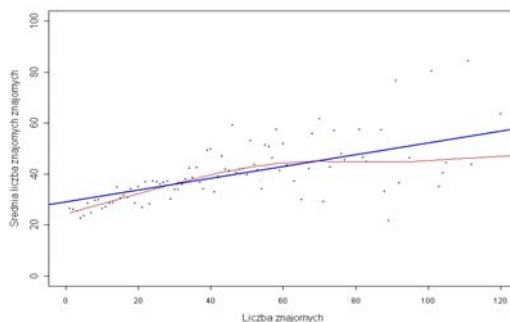
56

## Assortative mixing - degree

- ◆ Specjalnym przypadkiem assortative mixing jest:
  - degree correlation
- ◆ Sieci społeczne charakteryzują się własnością assortative mixing,
- ◆ Natomiast pozostałe rodzaje sieci (informacyjne, technologiczne, biologiczne) wykazują disassortative mixing (Newman, 2002).

57

## Zależność liczby relacji



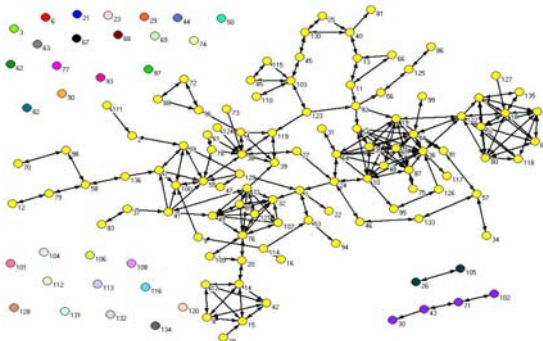
58

## Mixing patterns - wyniki

- ◆ Współczynnik korelacji Pearsona jest dobrą miarą zależności.
  - dodatni dla sieci z assortative mixing
  - ujemny dla sieci z disassortative mixing
- ◆ Zależność relacji od wieku,
  - $R^2=0,65$
- ◆ Zależność od liczby znajomych
  - $R^2=0,12$

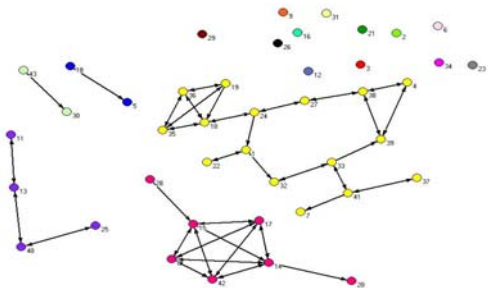
59

## Jądro sieci (250+ relacji)



60

## Jądro sieci (300+ relacji)



61

## Szukanie w sieci

- ◆ Modele małych światów nie wyjaśniają możliwości szukania w sieci
- ◆ Jon Kleinberg
  - Zależność relacji od odległości.
- ◆ Mark Newman i inni
  - Szukanie uwzględniające własności jednostek

62

## Rodzaje sieci

- ◆ Sieci:
  - Społeczne
  - Technologiczne
  - Informacyjne
  - Biologiczne
- ◆ Różne rodzaje sieci wykazują bardzo podobne własności

63

## Własności złożonych sieci

- ◆ Lokalna gęstość (clustering)
- ◆ Wielki komponent
- ◆ Krótkie ścieżki
- ◆ Potęgowy/ wykładniczy rozkład liczby relacji
- ◆ Mixing patterns
  - assortative mixing

64

## Konsekwencje struktury sieci

- ◆ Odporność sieci
- ◆ Dyfuzja informacji i innowacji
- ◆ Wpływ społeczny

65

## Odporność sieci

- ◆ Odporność na awarie:
  - losowe „awarie” węzłów.
- ◆ Odporność na atak:
  - eliminacja kluczowych (największych) węzłów.
- ◆ Znaczenie usuwania węzłów dla:
  - liczby komponentów i rozmiaru największego komponentu;
  - długości krótkich ścieżek;
  - efektywności sieci;

66

## Odporność sieci – wyniki

- ◆ Znaczenie różnych własności sieci dla jej odporności:
  - Sieci małych światów są odporne na awarie.
  - Sieci scale-free są odporne na awarie i niezwykle podatne na ataki.
  - Sieci wykazujące assortative mixing są bardziej odporne na ataki niż sieci z disassortative mixing.
- ◆ Konsekwencje

67

- ◆ Sieci, które na ogół chcielibyśmy przerwać,  
takie jak sieci społeczne rozprzestrzeniające chorobę (lub sieci terrorystów), mają pozytywną korelację wielkości sąsiadujących węzłów, a tym samym są bardziej odporne na ataki i izolowanie węzłów posiadających najwięcej kontaktów.
- ◆ Jednocześnie sieci, które chcielibyśmy chronić,  
na przykład sieci technologiczne takie jak Internet, mają korelację ujemną i są na takie ataki niezwykle podatne.

68

## Dyfuzja informacji

- ◆ Teoria perkolacji
  - Dwa istotne w epidemiologii i teorii perkolacji parametry to podatność na chorobę (*susceptibility*), czyli prawdopodobieństwo, że jednostka wystawiona na chorobę zarazi się nią i przekazywalność (*transmissibility*) - prawdopodobieństwo, że kontakt pomiędzy chorym a jednostką podatną na zarażenie zakończy się zarażeniem tej drugiej.
- ◆ Rozprzestrzenianie się informacji może zachodzić niezwykle szybko:
  - Bardzo krótkie ścieżki istnieją także w sieciach komunikacji krótkookresowej.

69

- ◆ Obieg informacji w sieciach, w których odległości są nieduże jest bardzo przyspieszony.
- ◆ Istnienie wielkiego komponentu w sieci relacji świadczy o możliwości bardzo powszechnego rozprzestrzenienia się informacji. Jego występowanie w sieci krótkookresowej komunikacji oznacza, że to rozpowszechnienie informacji może nastąpić niezwykle szybko.

70

## Wpływ społeczny

- ◆ Występowanie wielkiego komponentu w sieciach mocnych i długotrwałych relacji oznacza, że możliwe jest nie tylko rozprzestrzenianie informacji, ale również upowszechnianie norm, poglądów i opinii, a więc procesy mające charakter wpływu społecznego.
- ◆ Rozprzestrzenianie się idei wymaga równowagi pomiędzy zamkniętością grupy a połączeniami pomiędzy grupami
  - ochrona przed zmianą zdania
  - umożliwienie rozprzestrzeniania idei

71

- ◆ Mała, zamknięta grupa może dobrze utrzymywać swoje odrębne poglądy, jednak ma bardzo małą szansę na ich upowszechnienie. Większa otwarta społeczność jest na ogół dużo bardziej odporna na przyjęcie jakiejś jednej obowiązującej normy, ale z drugiej strony jej wprowadzenie może łatwo doprowadzić do jej bardzo szybkiego rozpowszechnienia się również poza tę społeczność.

72