Modelowanie i analiza sieci złożonych

III. Cechy sieci rzeczywistych i wizualizacja grafów.

Grzegorz Siudem

Politechnika Warszawska

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

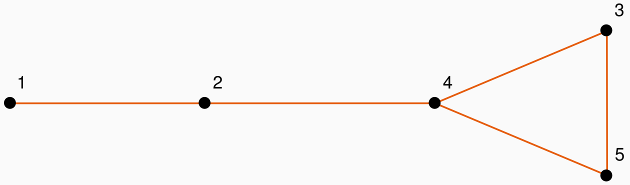
Przed zajęciami

Przypomnij sobie – stopień wierzchołka

Stopień wierzchołka (sieci nieskierowane)

Liczba krawędzi podłączonych do wierzchołka

$$k = \{1, 2, 2, 3, 2\},$$



Przypomnij sobie – rozkłady o grubych ogonach

Powtórka z probabilistyki i statystyki:

Przypomnij sobie – rozkłady o grubych ogonach

Powtórka z probabilistyki i statystyki:

- Czy rozkład prawdopodobieństwa może nie mieć wartości oczekiwanej? Wariancji?

Przypomnij sobie – rozkłady o grubych ogonach

Powtórka z probabilistyki i statystyki:

- Czy rozkład prawdopodobieństwa może nie mieć wartości oczekiwanej? Wariancji?
- Jakie są przykłady takich rozkładów?

Przypomnij sobie – rozkłady o grubych ogonach

Powtórka z probabilistyki i statystyki:

- Czy rozkład prawdopodobieństwa może nie mieć wartości oczekiwanej? Wariancji?
- Jakie są przykłady takich rozkładów?
- Co wiadomo o nośniku rozkładu, który nie ma skończonych momentów?

Przypomnij sobie – rozkłady o grubych ogonach

Powtórka z probabilistyki i statystyki:

- Czy rozkład prawdopodobieństwa może nie mieć wartości oczekiwanej? Wariancji?
- Jakie są przykłady takich rozkładów?
- Co wiadomo o nośniku rozkładu, który nie ma skończonych momentów?
- Jaką interpretację ma średnia (dobry estymator wartości oczekiwanej) gdy ta wartość oczekiwana nie istnieje? Do jakiej wartości zbiega wraz ze zwiększaniem próbki?

Wykład

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

- Rozmiar (małe i duże).

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

- Rozmiar (małe i duże).
- Planarność (por. wizualizacja).

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

- Rozmiar (małe i duże).
- Planarność (por. wizualizacja).
- Stopień regularności (choć dominują mniej regularne).

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

- Rozmiar (małe i duże).
- Planarność (por. wizualizacja).
- Stopień regularności (choć dominują mniej regularne).
- Typ grafu.

Jakie są rzeczywiste sieci złożone?

Co *nie wyróżnia* sieci rzeczywistych?

- Rozmiar (małe i duże).
- Planarność (por. wizualizacja).
- Stopień regularności (choć dominują mniej regularne).
- Typ grafu.
- ...

Przykłady?

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach
- Zjawisko małych światów.

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach
- Zjawisko małych światów.
- Korelacje pomiędzy wierzchołkami.

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach
- Zjawisko małych światów.
- Korelacje pomiędzy wierzchołkami.
- Obecność hierarchii (często).

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach
- Zjawisko małych światów.
- Korelacje pomiędzy wierzchołkami.
- Obecność hierarchii (często).
- Samopodobieństwo (czasami).

Jakie są sieci rzeczywiste?

Co wyróżnia sieci rzeczywiste?

- Rozkłady stopni wierzchołków o grubych ogonach
- Zjawisko małych światów.
- Korelacje pomiędzy wierzchołkami.
- Obecność hierarchii (często).
- Samopodobieństwo (czasami).

Uwaga!

Powyższe obserwacje są wynikami empirycznymi, a nie matematycznymi. Nie jest tak, **każda** sieć złożona posiada **każdą** z powyższych cech.

Co to są grube ogony?

Przypomnienie z probabilistyki

$$\mathbb{E}X^p = \int_{-\infty}^{\infty} x^p f(x) dx = \infty^\star$$

analogicznie dla rozkładów dyskretnych

$$\mathbb{E}X^p = \sum_{k=0}^{\infty} k^p P(k) = \infty^\star$$

Co to są grube ogony?

Przypomnienie z probabilistyki

$$\mathbb{E}X^p = \int_{-\infty}^{\infty} x^p f(x) dx = \infty^\star$$

analogicznie dla rozkładów dyskretnych

$$\mathbb{E}X^p = \sum_{k=0}^{\infty} k^p P(k) = \infty^\star$$

Uwaga!

Warunkiem koniecznym rozbiegania całek jest nieograniczony nośnik gęstości (funkcji masy p-stwa).

Co to są grube ogony?

Przypomnienie z probabilistyki

$$\mathbb{E}X^p = \int_{-\infty}^{\infty} x^p f(x) dx = \infty^\star$$

analogicznie dla rozkładów dyskretnych

$$\mathbb{E}X^p = \sum_{k=0}^{\infty} k^p P(k) = \infty^\star$$

Uwaga!

Warunkiem koniecznym rozbiegania całek jest nieograniczony nośnik gęstości (funkcji masy p-stwa).

Co jednak z sieciami rzeczywistymi?

Skończone czy nieskończone?

Dygresja o matematycznej precyzji

Jak sobie poradzić z tym problemem?

Typowo uznajemy, że rozkład stopni wierzchołków ma gruby ogon, jeśli odpowiednie całki (sumy) rozbiegają w granicy $N \to \infty$.

Dygresja o matematycznej precyzji

Jak sobie poradzić z tym problemem?

Typowo uznajemy, że rozkład stopni wierzchołków ma gruby ogon, jeśli odpowiednie całki (sumy) rozbiegają w granicy $N \to \infty$.

Przecież to niezgodne z matematyką! :(

Dygresja o matematycznej precyzji

Jak sobie poradzić z tym problemem?

Typowo uznajemy, że rozkład stopni wierzchołków ma gruby ogon, jeśli odpowiednie całki (sumy) rozbiegają w granicy $N \to \infty$.

Przecież to niezgodne z matematyką! :(

Zła wiadomość:

Takim stopniem precyzji cechuje się prawie cała sieciologia.

Dygresja o matematycznej precyzji

Jak sobie poradzić z tym problemem?

Typowo uznajemy, że rozkład stopni wierzchołków ma gruby ogon, jeśli odpowiednie całki (sumy) rozbiegają w granicy $N \to \infty$.

Przecież to niezgodne z matematyką! :(

Zła wiadomość:

Takim stopniem precyzji cechuje się prawie cała sieciologia.

Dobra wiadomość:

Jest to bardzo skuteczne podejście.

Dygresja o matematycznej precyzji

Jak sobie poradzić z tym problemem?

Typowo uznajemy, że rozkład stopni wierzchołków ma gruby ogon, jeśli odpowiednie całki (sumy) rozbiegają w granicy $N \to \infty$.

Przecież to niezgodne z matematyką! :(

Zła wiadomość:

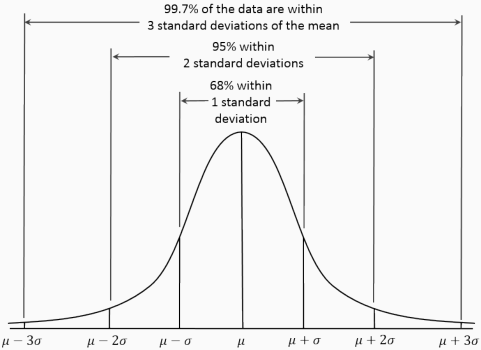
Takim stopniem precyzji cechuje się prawie cała sieciologia.

Dobra wiadomość:

Jest to bardzo skuteczne podejście.

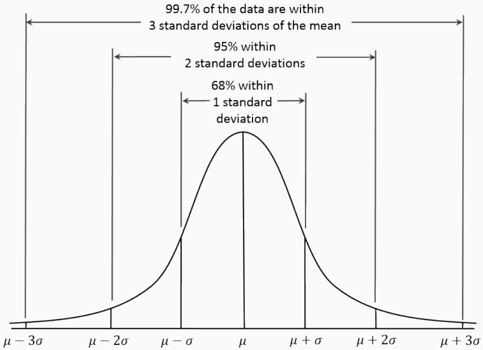
Bardziej matematyczno-precyzyjni będziemy na Wykładzie 7.

Czym skutkują grube ogony?



Źródło: wikipednia

Czym skutkują grube ogony?

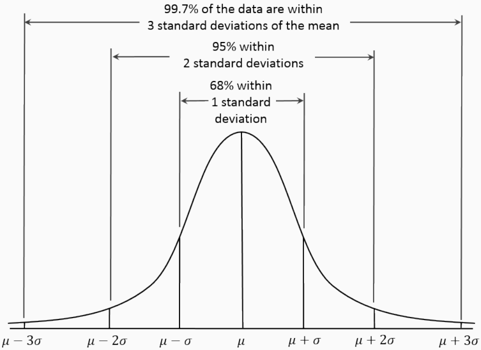


Źródło: wikipednia

Reguła Pareto (80/20)

Jaka część zasobów należy do jakiej części populacji?

Czym skutkują grube ogony?



Źródło: wikipednia

Reguła Pareto (80/20)

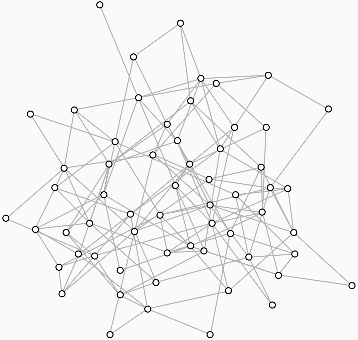
Jaka część zasobów należy do jakiej części populacji?

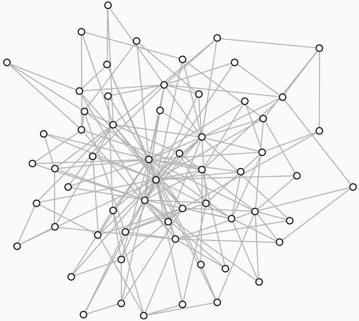
Czy to prawda dla dowolnych rozkładów?

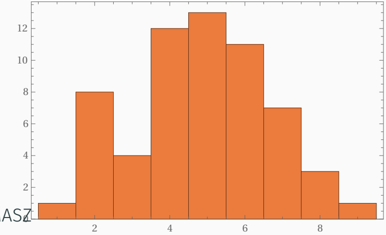
Sprawdzimy podczas projektu!

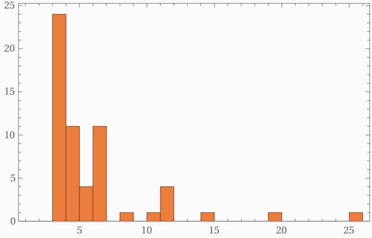
Dlaczego sieci potęgowe są bezskalowe?

Uzasadnienie empiryczne









Dlaczego sieci potęgowe są bezskalowe?

Uzasadnienie empiryczne

Zajmiemy się nim w trakcie części projekt!

Dlaczego sieci potęgowe są bezskalowe?

Uzasadnienie matematyczne

Rozkładami o grubych ogonach nazywamy takie, dla których nie istnieje wartość oczekiwana lub któryś z wyższych momentów.

$$\mathbb{E}X^k = \int f(x)x^k dx = \infty^\star$$

Dlaczego sieci potęgowe są bezskalowe?

Uzasadnienie matematyczne

Rozkładami o grubych ogonach nazywamy takie, dla których nie istnieje wartość oczekiwana lub któryś z wyższych momentów.

$$\mathbb{E}X^k = \int f(x)x^k dx = \infty^\star$$

Załóżmy, że nie istnieje drugi moment

$$\mathbb{E}X^2 = \infty,$$

a zatem

$$\mathbb{E}\left[X - \mathbb{E}X\right]^2 = \infty,$$

Dlaczego sieci potęgowe są bezskalowe?

Uzasadnienie matematyczne

Rozkładami o grubych ogonach nazywamy takie, dla których nie istnieje wartość oczekiwana lub któryś z wyższych momentów.

$$\mathbb{E}X^k = \int f(x)x^k dx = \infty^\star$$

Załóżmy, że nie istnieje drugi moment

$$\mathbb{E}X^2 = \infty,$$

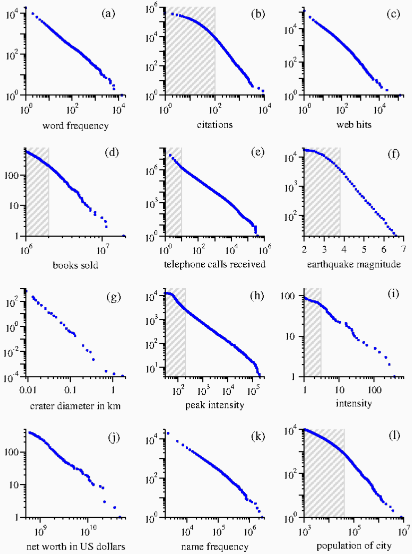
a zatem

$$\mathbb{E}\left[X - \mathbb{E}X\right]^2 = \infty,$$

Zatem nie mamy skali!

A jak to interpretować w przypadku skończonym?

Sieci o rozkładach potęgowych



M.E.J. Newman, *Power laws, Pareto distributions and Zipf's law*, Contemporary Physics, **46**, 323-351 (2005).

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,
- *rich get richer rule,*

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,
- *rich get richer rule*,
- magnetyzm sławy,

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,
- *rich get richer rule*,
- magnetyzm sławy,
- reguła preferencyjnego dołączania,

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,
- *rich get richer rule,*
- magnetyzm sławy,
- reguła preferencyjnego dołączania,
- zasada pieniądz robi pieniądz,

Reguła św. Mateusza

Albowiem wszelkiemu mającemu będzie dano, i obfitować będzie, a temu, który nie ma, i to, co się zda mieć, będzie wzięto od niego.

Mt 25:29.

- Reguła św. Mateusza,
- *rich get richer rule,*
- magnetyzm sławy,
- reguła preferencyjnego dołączania,
- zasada pieniądz robi pieniądz,
- ...

M. Perc, J. R. Soc. Interface, **11**, (2014).

Czy rozkład wystarczy?

Pytanie:

Należy rozstrzygnąć (udowodnić lub znaleźć kontrprzykład) czy rozkład stopni wierzchołków jednoznacznie charakteryzuje sieć/graf.

Czym mogą różnić się sieci o tym samym rozkładzie?

Asortatywność vs dysasortatywność

$$\mathcal{P}(k_i|k_j) = \frac{N\mathcal{P}(k_i, k_j)}{k_j\mathcal{P}(k_j)/\langle k \rangle}$$

Czym mogą różnić się sieci o tym samym rozkładzie?

Asortatywność vs dysasortatywność

$$\mathcal{P}(k_i|k_j) = \frac{N\mathcal{P}(k_i, k_j)}{k_j\mathcal{P}(k_j)/\langle k\rangle}$$

- Sieć dysasortatywna to taka, gdzie prawdopodobieństwo połączenia węzłów o bardzo różnym stopniu jest duże.

Czym mogą różnić się sieci o tym samym rozkładzie?

Asortatywność vs dysasortatywność

$$\mathcal{P}(k_i|k_j) = \frac{N\mathcal{P}(k_i, k_j)}{k_j\mathcal{P}(k_j)/\langle k\rangle}$$

- Sieć dysasortatywna to taka, gdzie prawdopodobieństwo połączenia węzłów o bardzo różnym stopniu jest duże.
- Sieć asortatywna to taka, gdzie prawdopodobieństwo połączenia węzłów o zbliżonym stopniu jest duże

Pytanie

Jakie są przykłady rzeczywiste takich sieci?

Czym mogą różnić się sieci o tym samym rozkładzie?

Korelacje

$\mathcal{P}(k_i, k_j)$ - prawdopodobieństwo, że losowo wybrana krawędź łączy węzły o stopniach k_i i k_j

$$\mathcal{R}(k_i, k_j) = \frac{\mathcal{P}(k_i, k_j)}{\mathcal{P}_u(k_i, k_j)},$$

a P_u odpowiada sieci nieskorelowanej o tym samym rozkładzie

Wyprowadzenie $\mathcal{P}_u$

Czym mogą różnić się sieci o tym samym rozkładzie?

Korelacje

$\mathcal{P}(k_i, k_j)$ - prawdopodobieństwo, że losowo wybrana krawędź łączy węzły o stopniach k_i i k_j

$$\mathcal{R}(k_i, k_j) = \frac{\mathcal{P}(k_i, k_j)}{\mathcal{P}_u(k_i, k_j)},$$

a P_u odpowiada sieci nieskorelowanej o tym samym rozkładzie

Wyprowadzenie $\mathcal{P}_u$

- Losowe przełączanie – ilustracja.

Czym mogą różnić się sieci o tym samym rozkładzie?

Korelacje

$\mathcal{P}(k_i, k_j)$ - prawdopodobieństwo, że losowo wybrana krawędź łączy węzły o stopniach k_i i k_j

$$\mathcal{R}(k_i, k_j) = \frac{\mathcal{P}(k_i, k_j)}{\mathcal{P}_u(k_i, k_j)},$$

a P_u odpowiada sieci nieskorelowanej o tym samym rozkładzie

Wyprowadzenie $\mathcal{P}_u$

- Losowe przełączanie – ilustracja.
- $\mathcal{P}_u(k_i, k_j) = \frac{k_i k_j \mathcal{P}(k_i)\mathcal{P}(k_j)}{\langle k \rangle^2}$

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Jak zmierzyć małość świata?

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

Ile wynosi średnia droga w wybranych modelach sieci?

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Jak zmierzyć małość świata?

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

Ile wynosi średnia droga w wybranych modelach sieci?

- dla grafów przypadkowych $\ell \sim \frac{\ln N}{\ln \langle k \rangle}$,

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Jak zmierzyć małość świata?

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

Ile wynosi średnia droga w wybranych modelach sieci?

- dla grafów przypadkowych $\ell \sim \frac{\ln N}{\ln \langle k \rangle}$,
- dla siatki kwadratowej $\ell \sim \sqrt{N}$,

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Jak zmierzyć małość świata?

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

Ile wynosi średnia droga w wybranych modelach sieci?

- dla grafów przypadkowych $\ell \sim \frac{\ln N}{\ln\langle k \rangle}$,
- dla siatki kwadratowej $\ell \sim \sqrt{N}$,
- dla sieci bezskalowych z $\alpha = 3$ $\ell \sim \frac{\ln N}{\ln \ln N}$,

Inne cechy sieci rzeczywistych – małe światy

Przypomnijmy:

Ile uścisków dłoni dzieli dwóch dowolnych mieszkańców Ziemi? (eksperyment Milgrama)

Jak zmierzyć małość świata?

$$\ell = \frac{1}{N(N-1)} \sum_{i \neq j} d(i,j)$$

Ile wynosi średnia droga w wybranych modelach sieci?

- dla grafów przypadkowych $\ell \sim \frac{\ln N}{\ln \langle k \rangle}$,
- dla siatki kwadratowej $\ell \sim \sqrt{N}$,
- dla sieci bezskalowych z $\alpha = 3$ $\ell \sim \frac{\ln N}{\ln \ln N}$,
- dla sieci bezskalowych z $\alpha \in (2,3)$ $\ell \sim \ln \ln N$.

Inne cechy wybranych sieci rzeczywistych

- fraktalne sieci dystrybucyjne,

Inne cechy wybranych sieci rzeczywistych

- fraktalne sieci dystrybucyjne,
- sieci hierarchiczne,

Inne cechy wybranych sieci rzeczywistych

- fraktalne sieci dystrybucyjne,
- sieci hierarchiczne,
- sieci ze strukturą społeczną.

Pytanie:

Znajdź rzeczywisty przykład powyższych sieci.

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,
- odpowiednie kąty pomiędzy krawędziami,

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,
- odpowiednie kąty pomiędzy krawędziami,
- minimalna liczba przecięć,

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,
- odpowiednie kąty pomiędzy krawędziami,
- minimalna liczba przecięć,
- symetrie!

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,
- odpowiednie kąty pomiędzy krawędziami,
- minimalna liczba przecięć,
- symetrie!
- jest *przyjemna dla oka*.

Uwaga

Optymalizacja powyższych cech (zwłaszcza ostatniej!) jest wymagająca algorytmicznie.

Metody wizualizacji grafow

Cechy dobrej wizualizacji:

- odpowiednia odległość pomiędzy wierzchołkami,
- odpowiednie kąty pomiędzy krawędziami,
- minimalna liczba przecięć,
- symetrie!
- jest *przyjemna dla oka*.

Uwaga

Optymalizacja powyższych cech (zwłaszcza ostatniej!) jest wymagająca algorytmicznie.

Wniosek:

Wizualizacja to w dużej mierze sztuka... lub wykorzystanie gotowych narzędzi.

Dla fanów R-a

`http://kateto.net/network-visualization`

Podsumowanie

Na następne zajęcia:

Przeczytaj

- M.E.J. Newman, *Power laws, Pareto distributions and Zipf's law*, Contemporary Physics, **46**, 323-351 (2005).
  i/lub item rozdziały 3.1-3.3 w A. Fronczak, P. Fronczak, *Świat sieci złożonych* PWN (2009).

Przeczytaj

Pobierz i przygotuj się do analizy wybranych danych o potęgowym rozkładzie.

Dziękuję za uwagę!

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