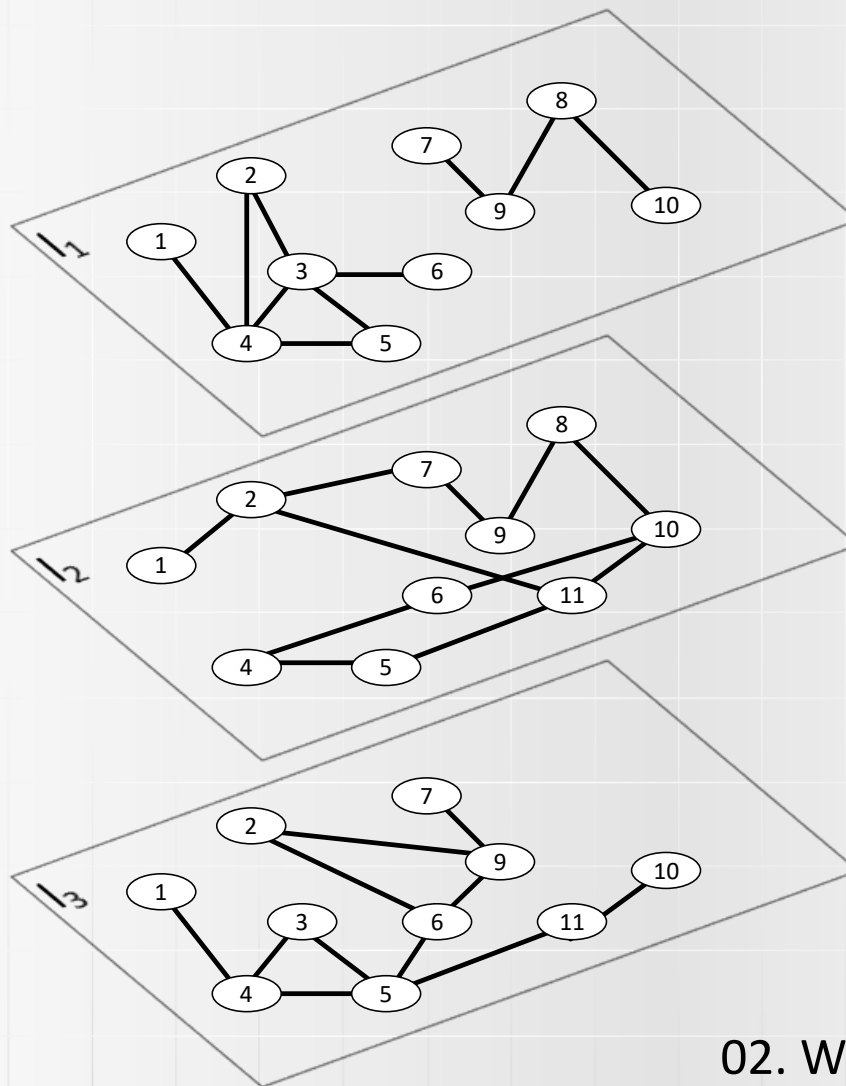




Politechnika
Wrocławska

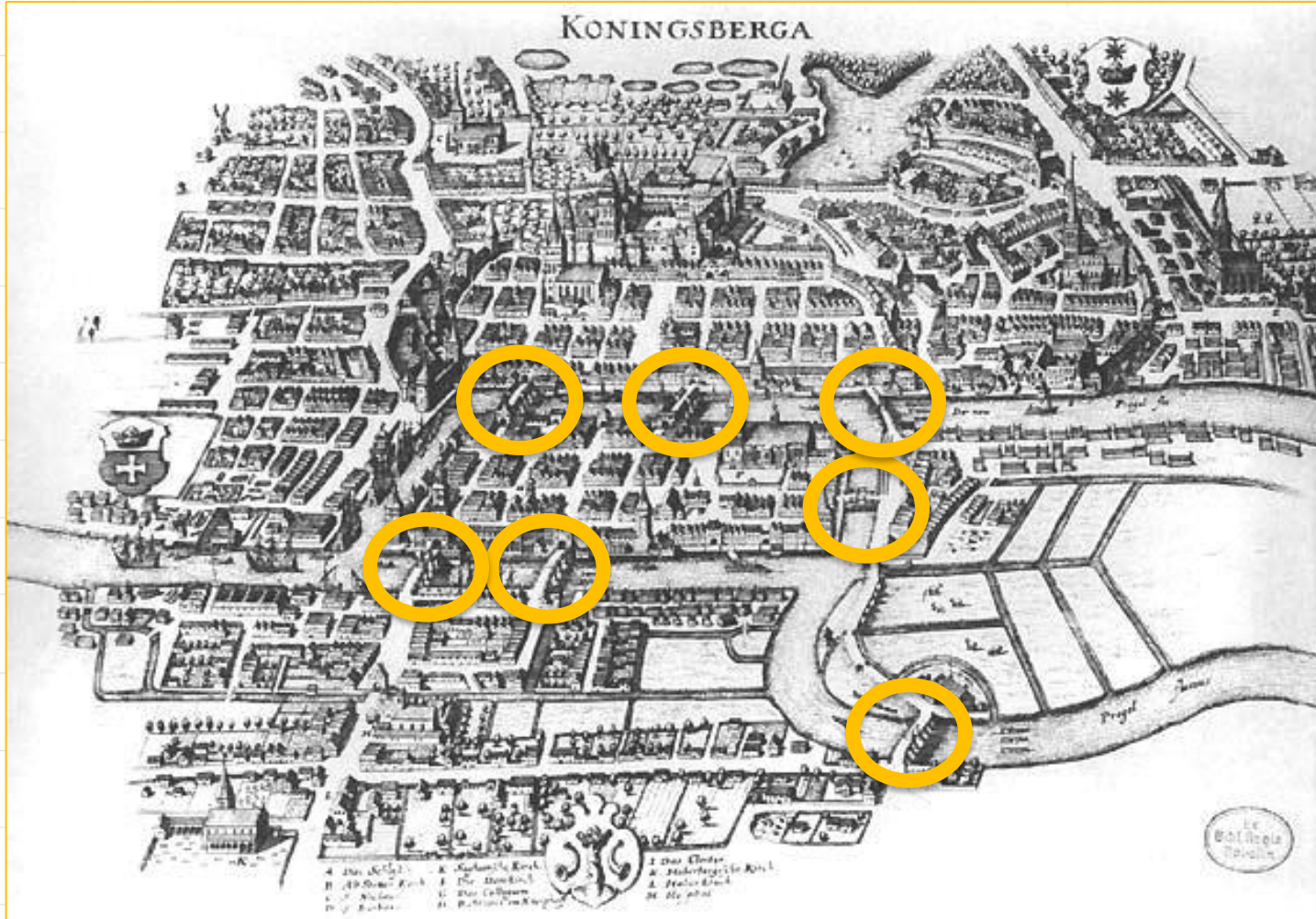
Sieci złożone



02. Wstęp do teorii grafów



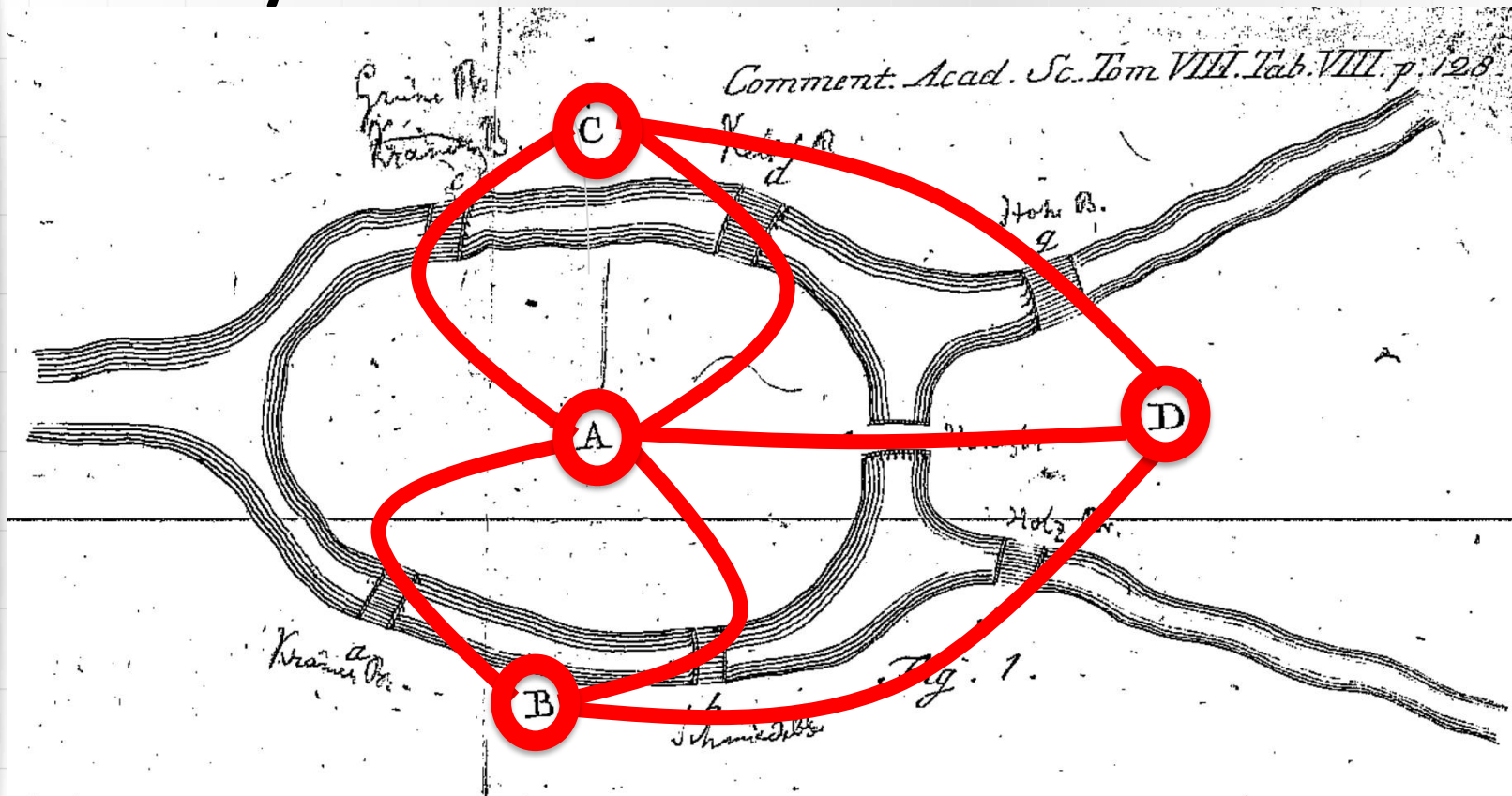
Mosty Królewca - 1736



Czy można przejść przez siedem mostów i nigdy nie przekroczyć tego samego mostu dwa razy?

Euler L., *Solutio problematis ad geometriam situs pertinentis* w *Commentarii academiae scientiarum Petropolitanae*, wolumen 8, strony 128-140 (łac.)

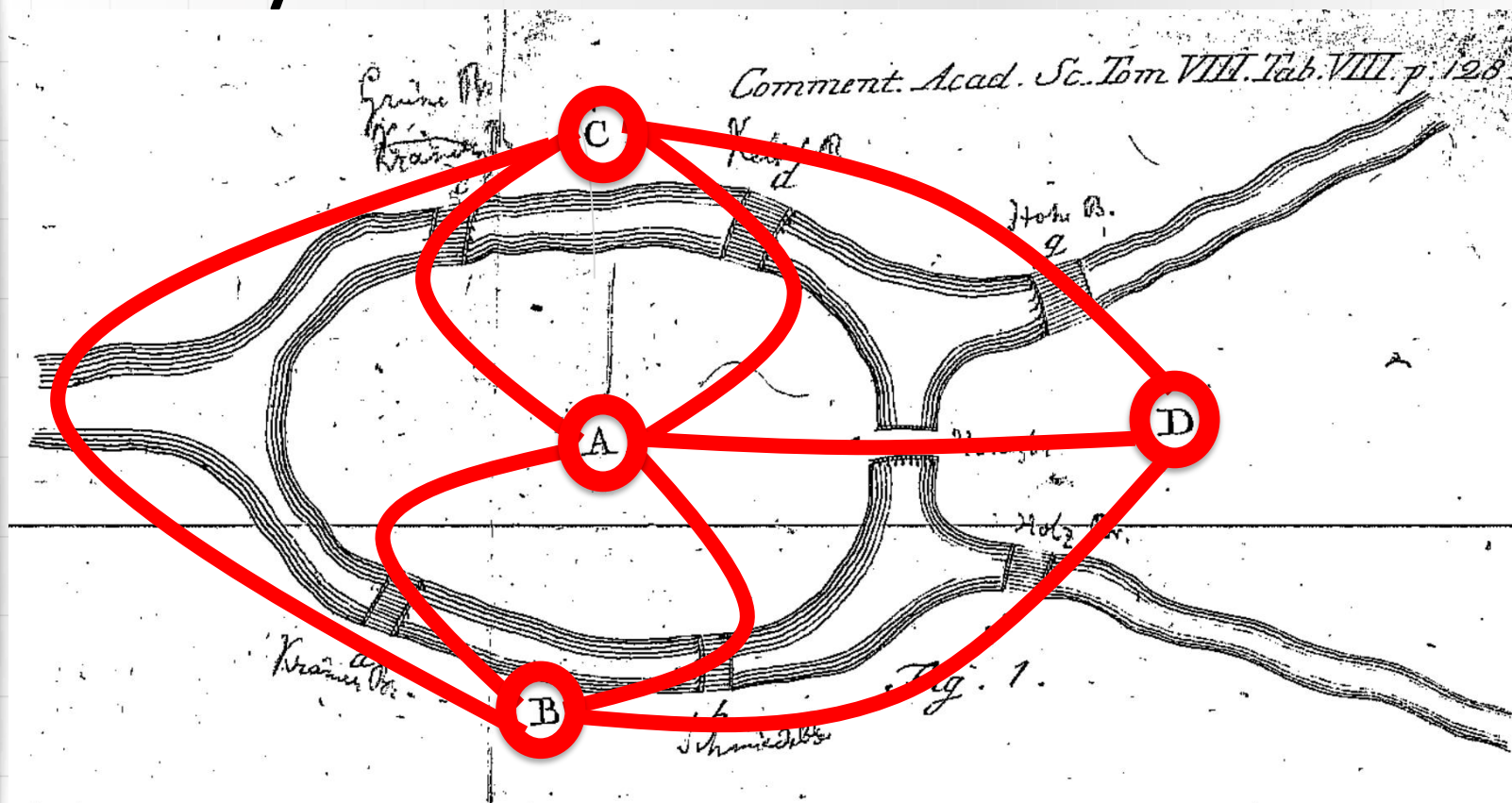
Mosty Królewca



1736: Rozwiązanie Eulera:

- (a) Jeśli graf ma **więcej niż dwa wierzchołki** o nieparzystym stopniu, taka ścieżka nie istnieje.
- (b) Jeśli graf jest połączonym i ma **nie więcej niż dwa wierzchołki** o nieparzystym stopniu, istnieje co najmniej jedna taka ścieżka.

Mosty Królewca



1875:

Władze miasta rozwiązały problem



LORD MAYOR OF KÖNIGSBERG, I WILL
REWARD YOU HANDSOMELY IF YOU
CONSTRUCT THIS BRIDGE BEFORE
MY FRIEND LEONHARD ARRIVES.



I TRIED TO USE A TIME MACHINE TO CHEAT ON MY ALGORITHMS
FINAL BY PREVENTING GRAPH THEORY FROM BEING INVENTED.



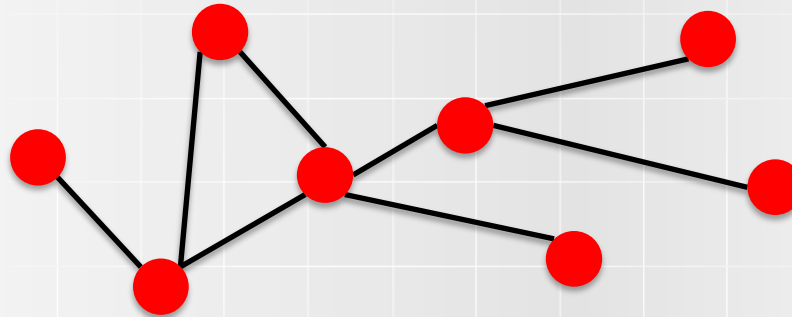
Politechnika
Wrocławska

Czy graf = sieć?



HR EXCELLENCE IN RESEARCH

Elementy złożonego systemu



- Komponenty sieci N (lub V) – węzły, wierzchołki (nodes, vertices)
- Interakcje, zależności pomiędzy komponentami L (lub E) – połączenia, krawędzie (links, edges)
- Cały system (N, L) – sieć (network), graf (graph)

Sieć czy graf?

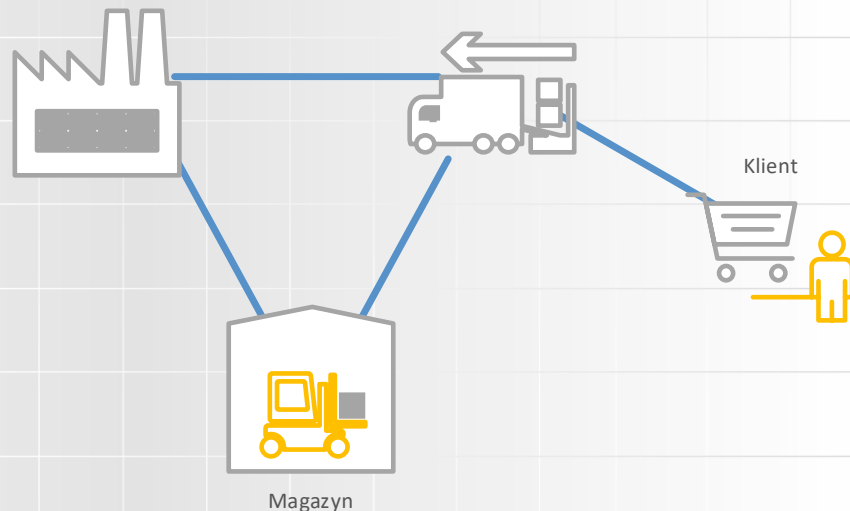
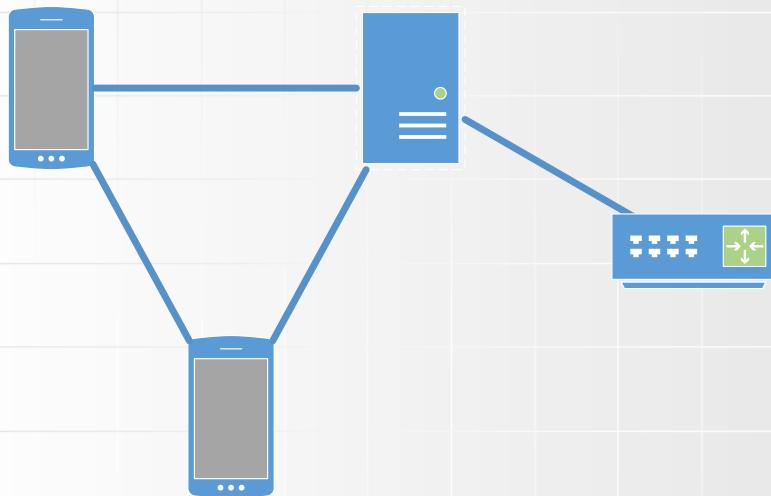
Sieć zazwyczaj odnosi się do rzeczywistego systemu

- *Sieć www*
- *Sieć społeczna*

Graf jest jedną z wielu możliwych matematycznych reprezentacji sieci

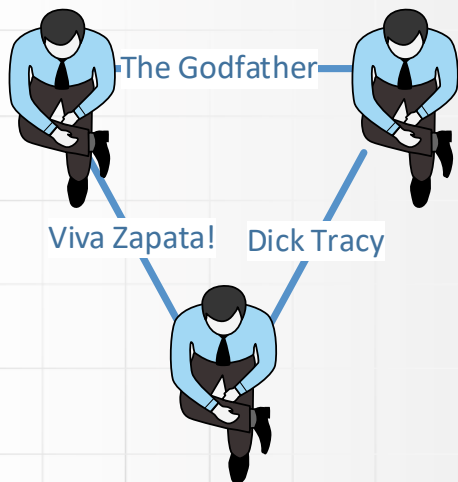
- Graf odnośników pomiędzy stronami www
- Graf relacji przyjaźni pomiędzy studentami

Wiele sieci jeden graf



Marion Brando

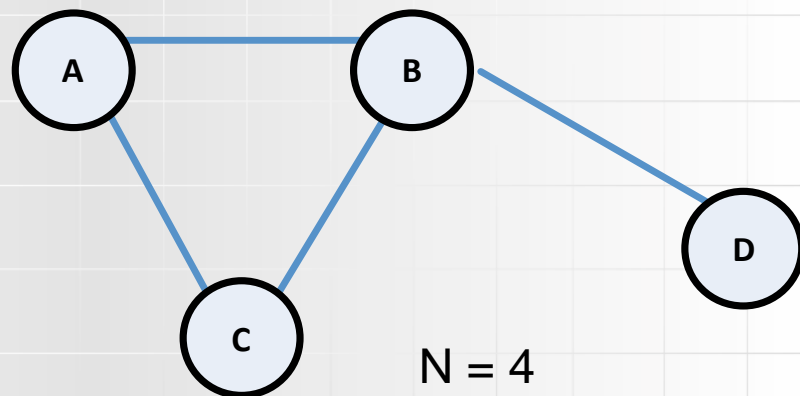
Al. Pacino



Scareface

Michelle Pfeiffer

Henry Silva



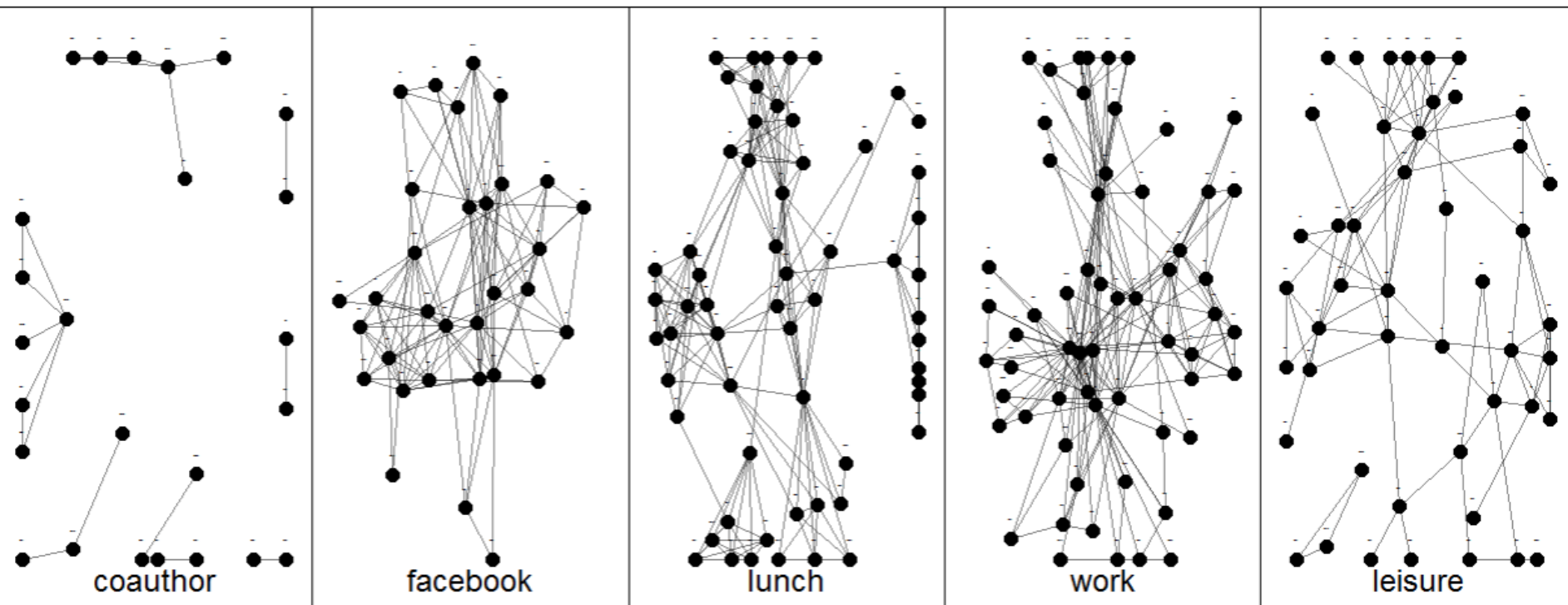
$N = 4$
 $L = 4$

Wiele sieci jeden system.

- Wybór właściwej reprezentacji sieci determinuje naszą zdolność do skutecznego wykorzystania teorii sieci.
- W niektórych przypadkach istnieje jednoznaczna reprezentacja. W innych przypadkach możemy mieć wiele różnych.
- Na przykład sposób, w jaki przypiszemy powiązania między osobami, określi charakter naszych badań.

Wybór właściwej reprezentacji

AUCS network



Wybór właściwej reprezentacji

Jeśli połączysz osoby na podstawie ich imienia
(wszyscy Piotrowie są ze sobą połączeni), co
możesz zbadać?

Niemniej dalej jest to sieć!



Politechnika
Wrocławska

Grafy nieskierowane i skierowane

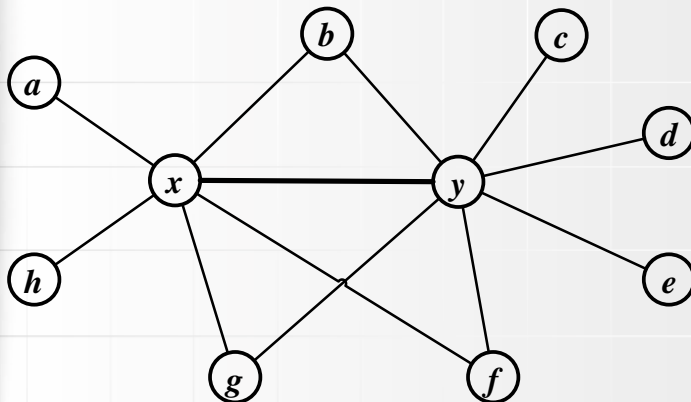


HR EXCELLENCE IN RESEARCH

Grafy nieskierowane i skierowane

Nieskierowane

Krawędzie: nieskierowane (relacja symetryczna)



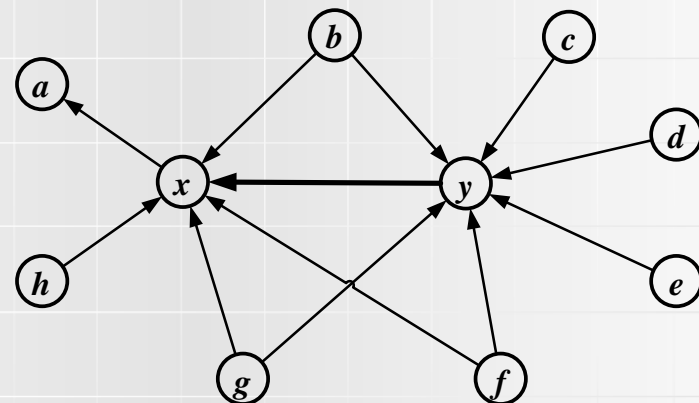
Przykłady

Sieć współpracy naukowej
Sieć aktorów
Interakcje pomiędzy białkami

Skierowane

Krawędzie: skierowane (*relacja niesymetryczna*)

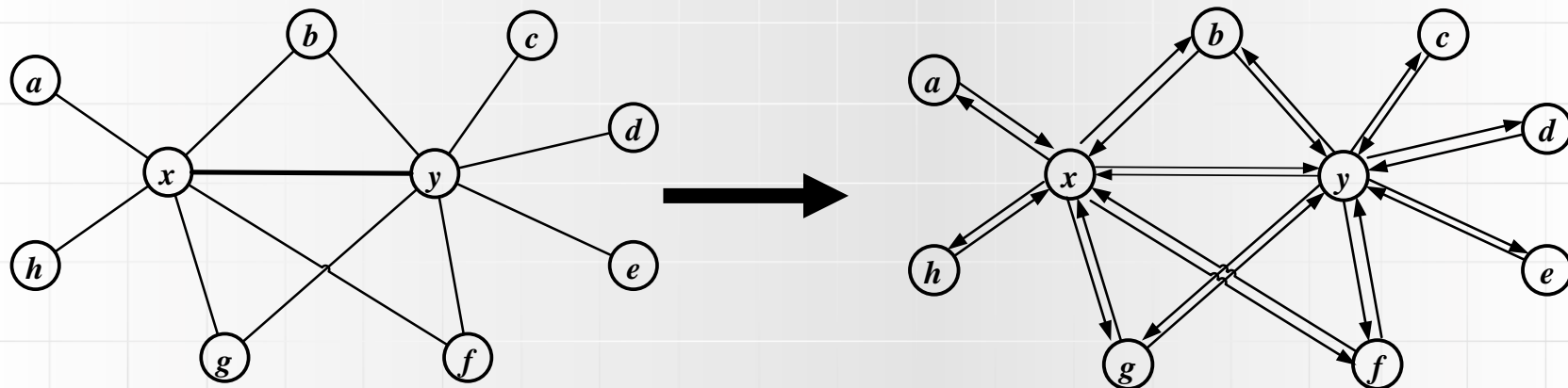
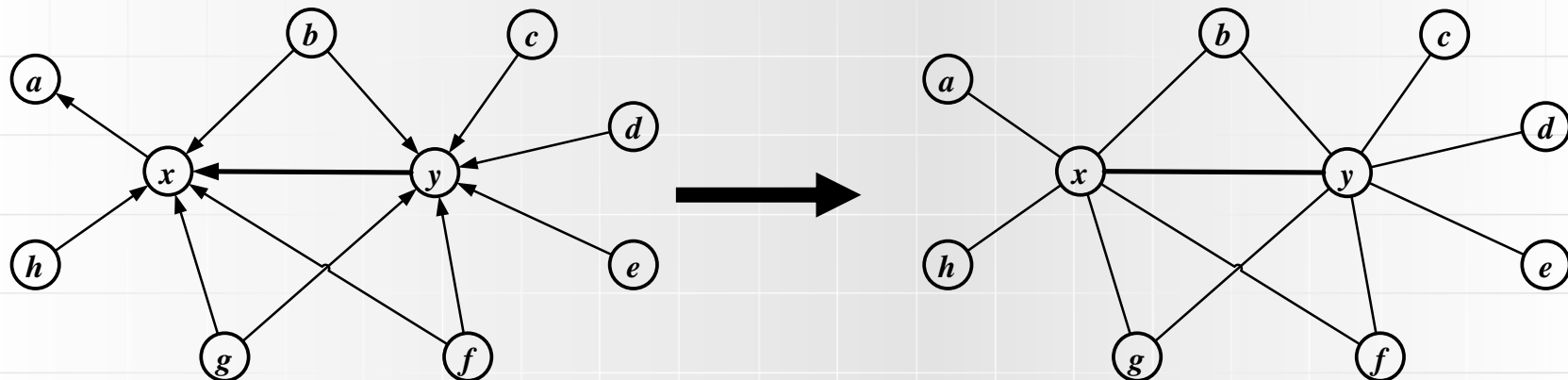
Graf skierowany (ang. Digraph, directed graph):



Przykłady

Odnośniki na stronach WWW
Rozmowy telefoniczne
Uczucia

Grafy nieskierowane i skierowane



Przykładowe sieci

Sieć	Wierzchołki	Krawędzie	Skierowana/nieskierowana	Liczba wierzchołków	Liczba Krawędzi	Średni stopień wężła
Internet						
WWW						
Sieć energetyczna						
Rozmowy telefoniczne						
Emaile						
Współpraca naukowa						
Sieć aktorów						
Sieć cytowań						
Metabolizm Ecoli						
Interakcje pomiędzy proteinami						

Przykładowe sieci

Sieć	Wierzchołki	Krawędzie	Skierowana/ni es kierowana	Liczba wierzchołków	Liczba Krawędzi	Średni stopień węzła
Internet	Rutery	Połączenia internetowe	Nieskierowana	192,244	609,066	6.34
WWW	Strony www	Linki	Skierowana	325,729	1,497,134	4.60
Sieć energetyczna	Elektrownie, transformatory	Kable	Nieskierowana	4,941	6,594	2.67
Rozmowy telefoniczne	Abonenci	Połączenia	Skierowana	36,595	91,826	2.51
Emaile	Adresy email	Emaile	Skierowana	57,194	103,731	1.81
Współpraca naukowa	Naukowcy	Współautorstwo	Nieskierowana	23,133	93,437	8.08
Sieć aktorów	Aktorzy	Wspólna gra w filmie	Nieskierowana	702,388	29,397,908	83.71
Sieć cytowań	Art. naukowe	Cytowania	Skierowana	449,673	4,689,479	10.43
Metabolizm Ecoli	Metabolity	Reakcje chemiczne	Skierowana	1,039	5,802	5.58
Interakcje pomiędzy proteinami	Proteiny (Białka)	Reakcje łączenia	Nieskierowana	2,018	2,930	2.90



Politechnika
Wrocławska

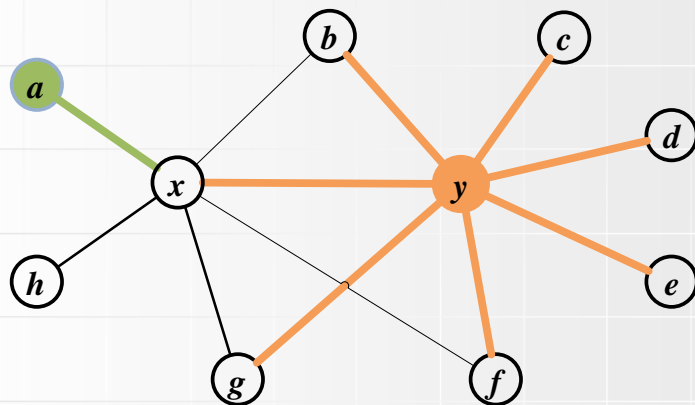
Stopień wężła (node degree)



HR EXCELLENCE IN RESEARCH

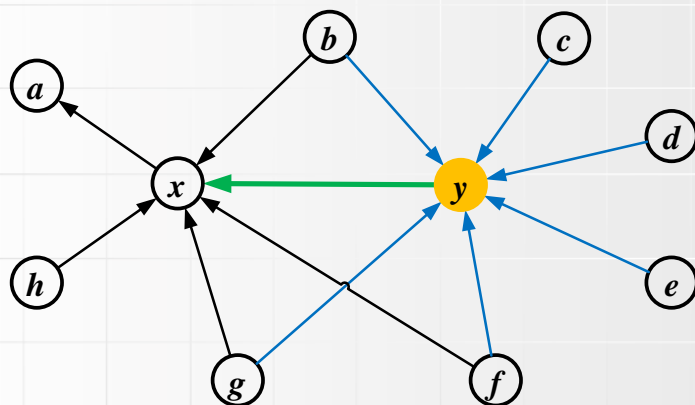
Stopień Węzła k_i

Nieskierowana



$$k_a = 1, k_y = 7$$

Skierowana

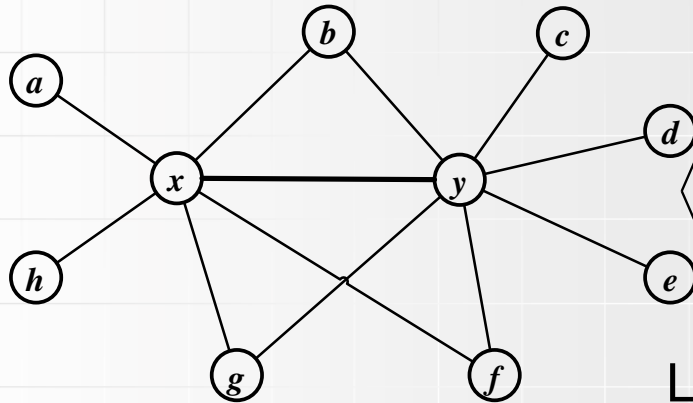


$$k_y^{in} = 6 \quad k_y^{out} = 1 \quad k_y = 7$$

Stopień węzła i : liczba krawędzi połączonych z węzłem i .

Średni stopień wężła $\langle k \rangle$

Nieskierowana

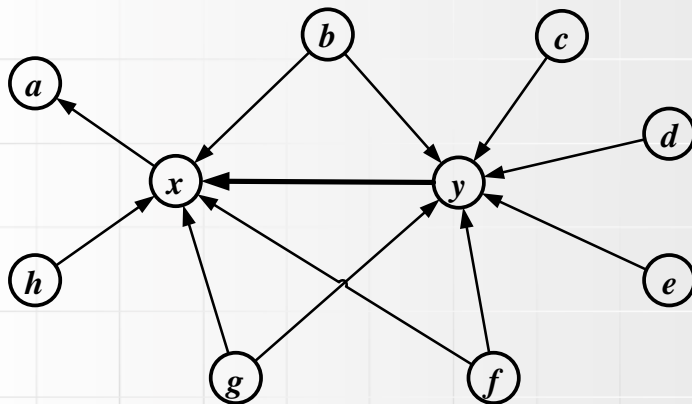


$$\langle k \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i \quad \langle k \rangle = \frac{2L}{N} = \frac{2 \cdot 12}{10}$$

L – liczba krawędzi w grafie

N – liczba wierzchołków w grafie

Skierowana

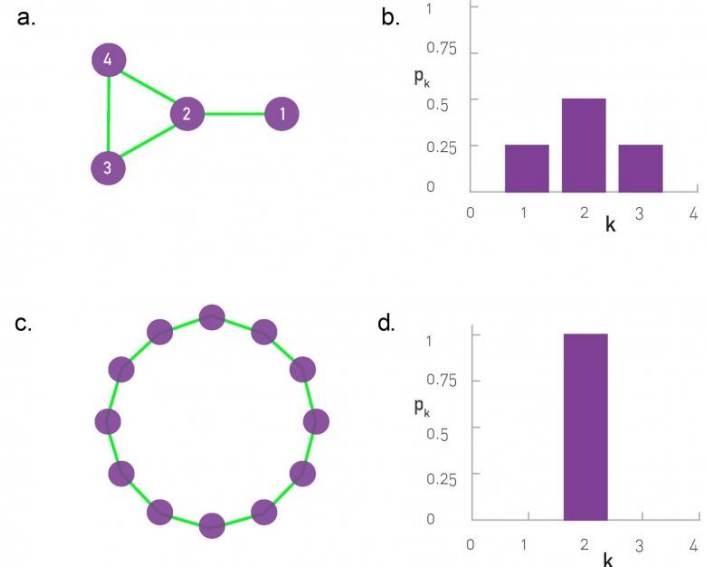


$$\langle k^{in} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{in}, \quad \langle k^{out} \rangle \equiv \frac{1}{N} \sum_{i=1}^N k_i^{out}, \quad \langle k^{in} \rangle = \langle k^{out} \rangle$$

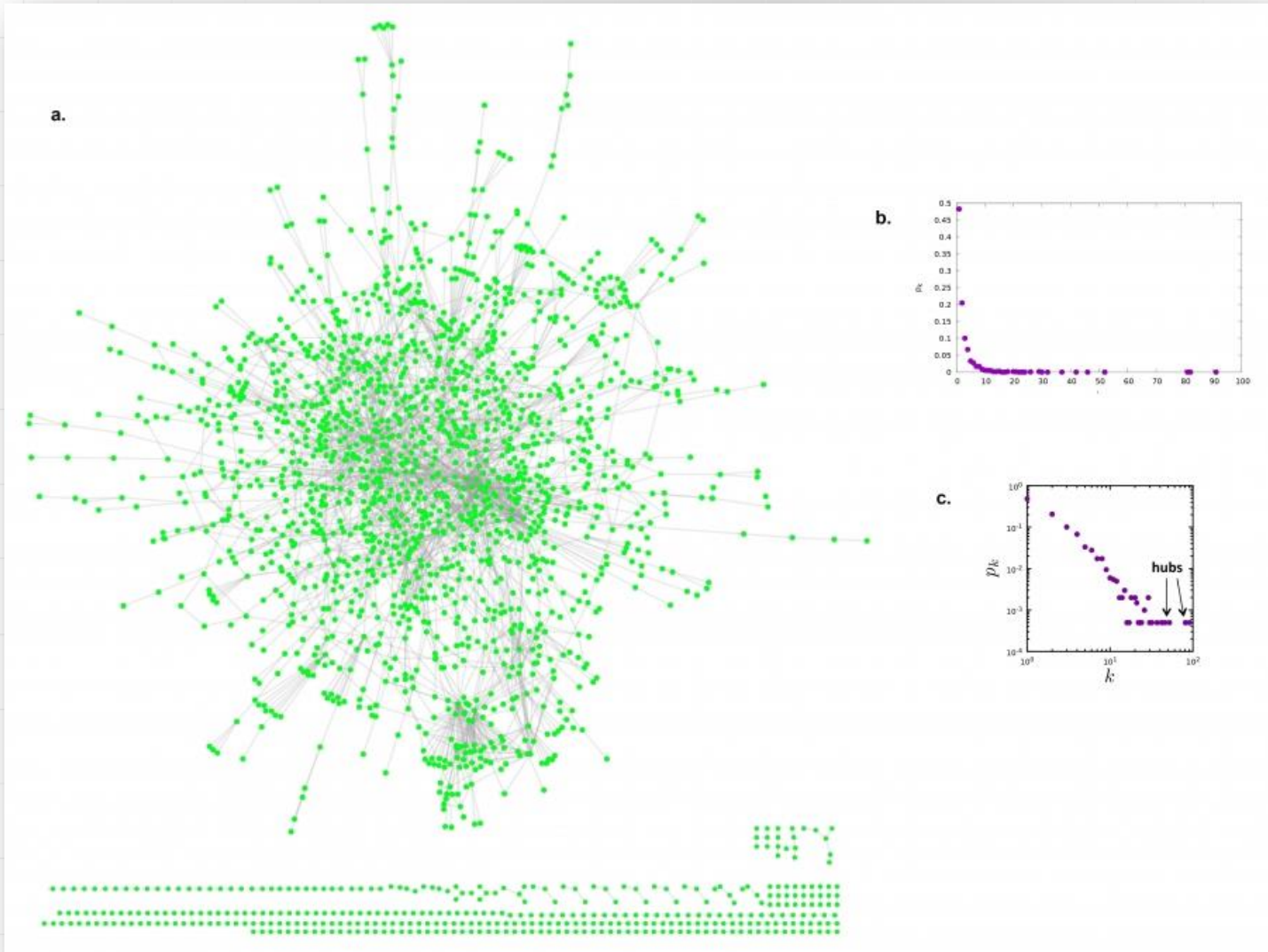
$$\langle k \rangle = \frac{L}{N} = \frac{12}{10}$$

Rozkład stopni wężła $P(k)/P_k$

- $P(k)$ jest to prawdopodobieństwo, że losowo wybrany wierzchołek w sieci ma stopień wężła k
- $P(k) = N_k / N$
- N_k – liczba wierzchołków o stopniu k



Rozkład stopni węzła





Politechnika
Wrocławska

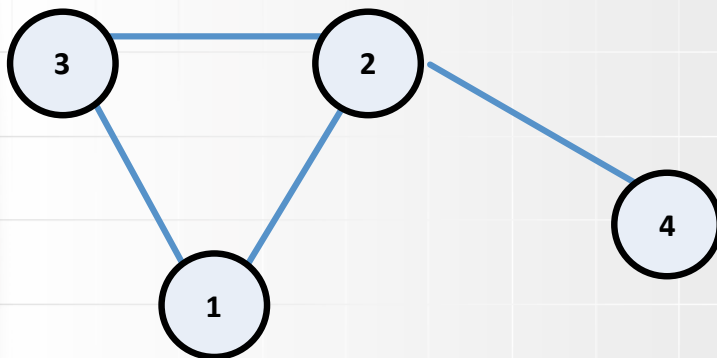
Macierz sąsiedztwa (Adjacency matrix)



HR EXCELLENCE IN RESEARCH

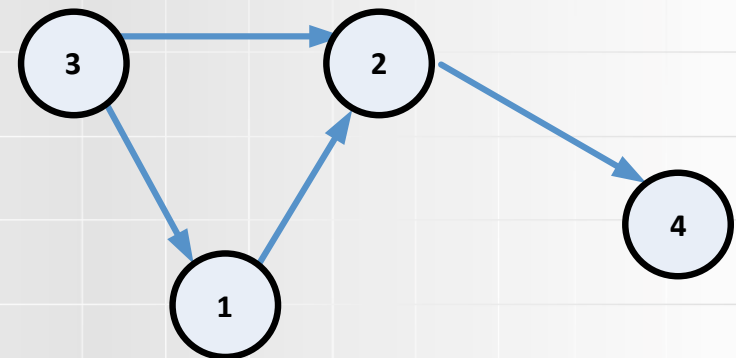
Macierz sąsiedztwa

Nieskierowane



0	1	1	0
1	0	1	1
1	1	0	0
0	1	0	0

Skierowane



0	0	1	0
1	0	1	0
0	0	0	0
0	1	0	0

$A_{ij} = 1$ jeżeli istnieje krawędź od i do j

$A_{ij} = 0$ jeżeli nie istnieje krawędź od i do j

Macierz sąsiedztwa

Nieskierowane

0	1	1	0
1	0	1	1
1	1	0	0
0	1	0	0

Skierowane

0	0	1	0
1	0	1	0
0	0	0	0
0	1	0	0

$$k_2 = \sum_{j=1}^4 A_{2j} = \sum_{i=1}^4 A_{i2} = 3 \quad k_2^{\text{in}} = \sum_{j=1}^4 A_{2j} = 2, \quad k_2^{\text{out}} = \sum_{i=1}^4 A_{i2} = 1$$

$$A_{ij} = A_{ji} \quad A_{ii} = 0 \quad \star$$

$$A_{ij} \neq A_{ji} \quad A_{ii} = 0$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij}$$

$$L = \sum_{i,j=1}^N A_{ij}$$

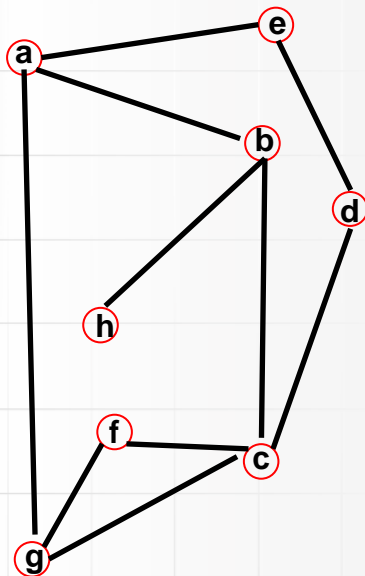
$$\langle k \rangle = \frac{2L}{N}$$

$$\langle k^{\text{in}} \rangle = \langle k^{\text{out}} \rangle = \frac{L}{N}$$

Macierz sąsiedztwa

	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	1	0	0
h	0	1	0	0	0	0	0	0

Macierz sąsiedztwa



	a	b	c	d	e	f	g	h
a	0	1	0	0	1	0	1	0
b	1	0	1	0	0	0	0	1
c	0	1	0	1	0	1	1	0
d	0	0	1	0	1	0	0	0
e	1	0	0	1	0	0	0	0
f	0	0	1	0	0	0	1	0
g	1	0	1	0	0	1	0	0
h	0	1	0	0	0	0	0	0



Politechnika
Wrocławska

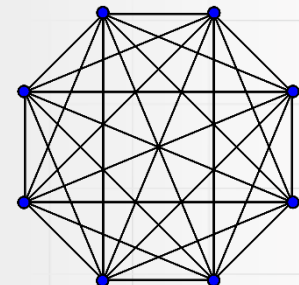
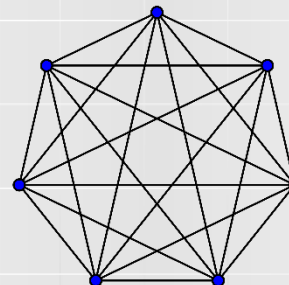
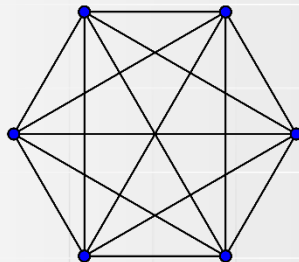
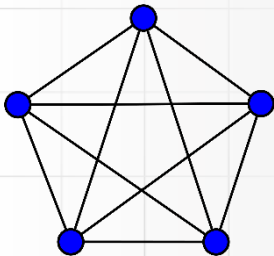
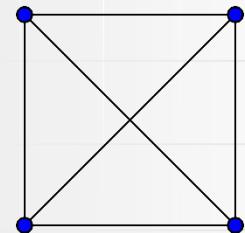
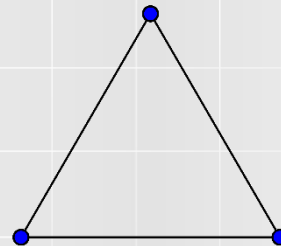
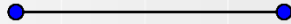
Rzeczywiste sieci są rzadkie



HR EXCELLENCE IN RESEARCH

Graf pełny

- Maksymalna liczba krawędzi w sieci z N wierzchołkami wynosi: $L_{max} = \frac{N(N-1)}{2}$
- Graf mający $L = L_{max}$ nazywamy grafem pełnym a jego średni stopień wężła wynosi $\langle k \rangle = N-1$



Rzeczywiste sieci są rzadkie

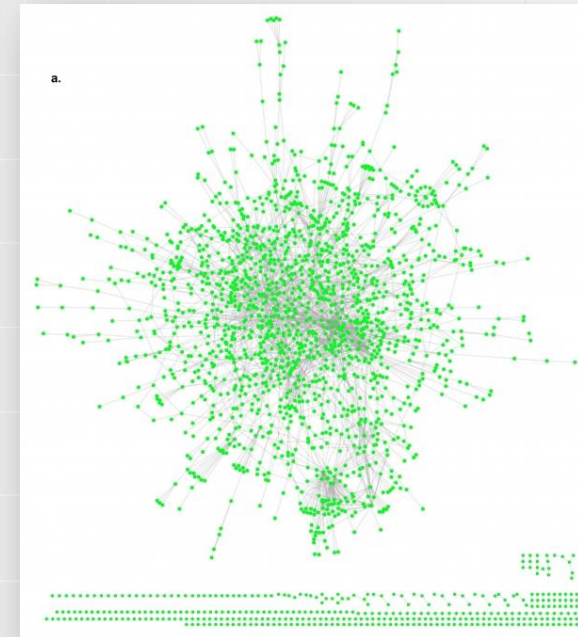
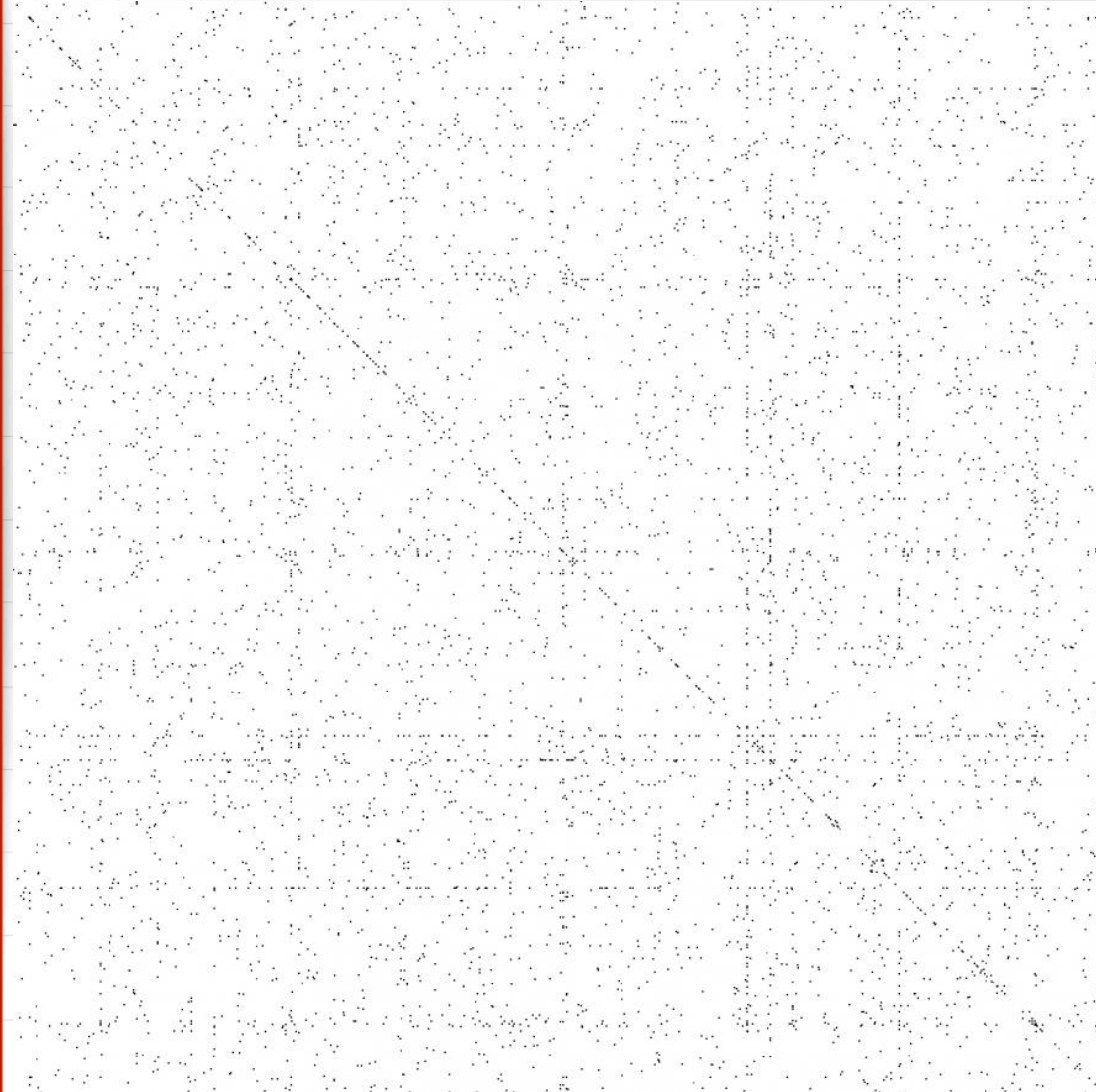
- Większość sieci obserwowanych w rzeczywistych systemach jest rzadka:

$$L \ll L_{\max}$$

$$\langle k \rangle \ll N-1$$

Sieć	Liczba wierzchołków	Liczba Krawędzi	Średni stopień wężła
Internet	192,244	609,066	6.34
WWW	325,729	1,497,134	4.60
Sieć energetyczna	4,941	6,594	2.67
Rozmowy telefoniczne	36,595	91,826	2.51
Email	57,194	103,731	1.81
Współpraca naukowa	23,133	93,437	8.08
Sieć aktorów	702,388	29,397,908	83.71
Sieć cytowań	449,673	4,689,479	10.43
Metabolizm Ecoli	1,039	5,802	5.58
Interakcje pomiędzy proteinami	2,018	2,930	2.90

Rzeczywiste sieci są rzadkie





Politechnika
Wrocławska

Sieć dwudzielna (Bipartite network)

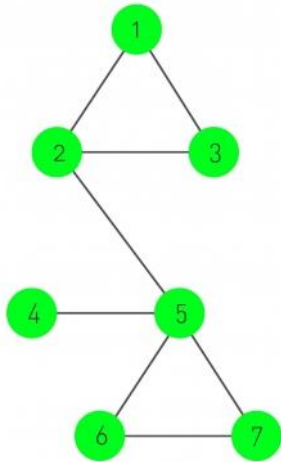


HR EXCELLENCE IN RESEARCH

Sieć dwudzielna

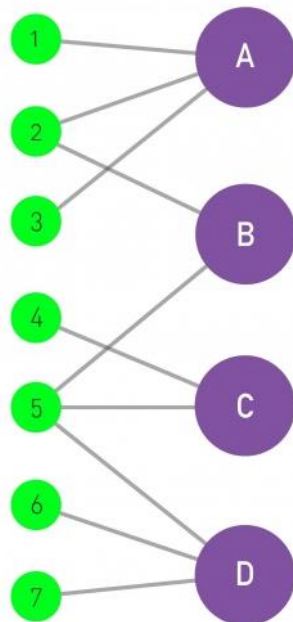
Graf dwudzielny jest to graf którego wierzchołki mogą być podzielone na dwa rozdzielne zbiory U i V tak że każda krawędź łączy wierzchołek ze zbioru U i wierzchołek ze zbioru V .

PROJECTION U U

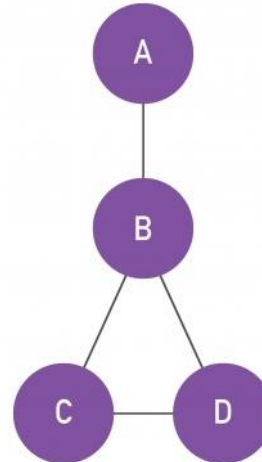


U

V



PROJECTION V V



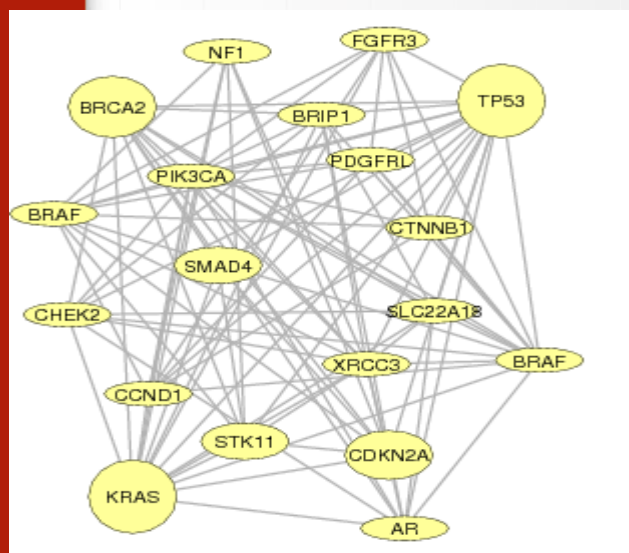
Przykłady:

Sieć aktorów

Sieć współautorstwa

Sieć chorób

Sieć genów – Sieć chorób

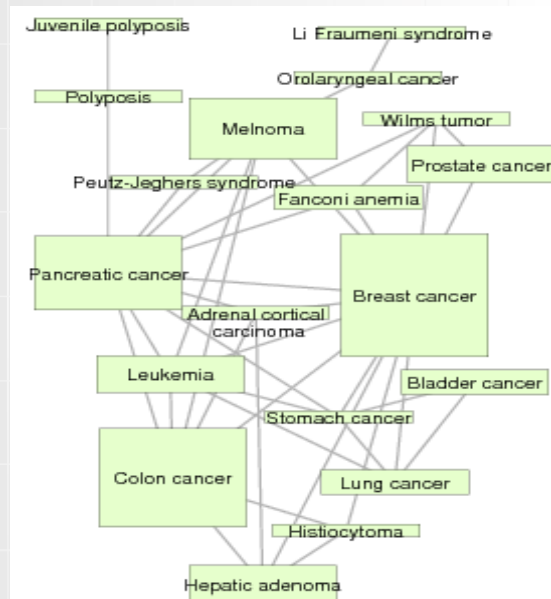
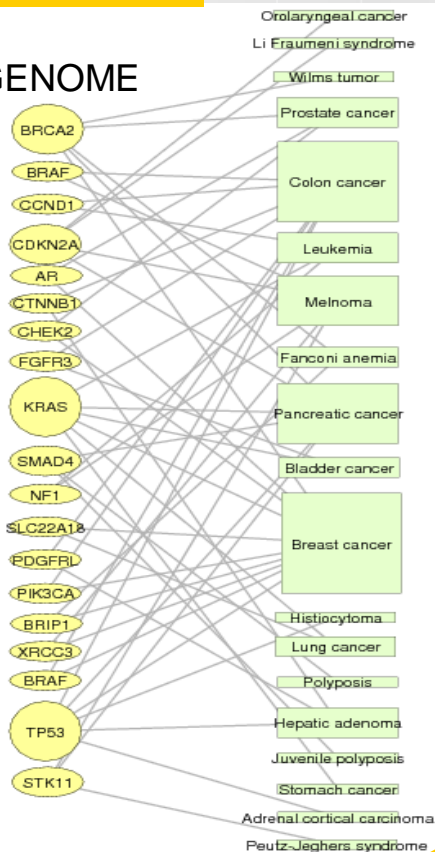


Sieć Genów

DISEASOME

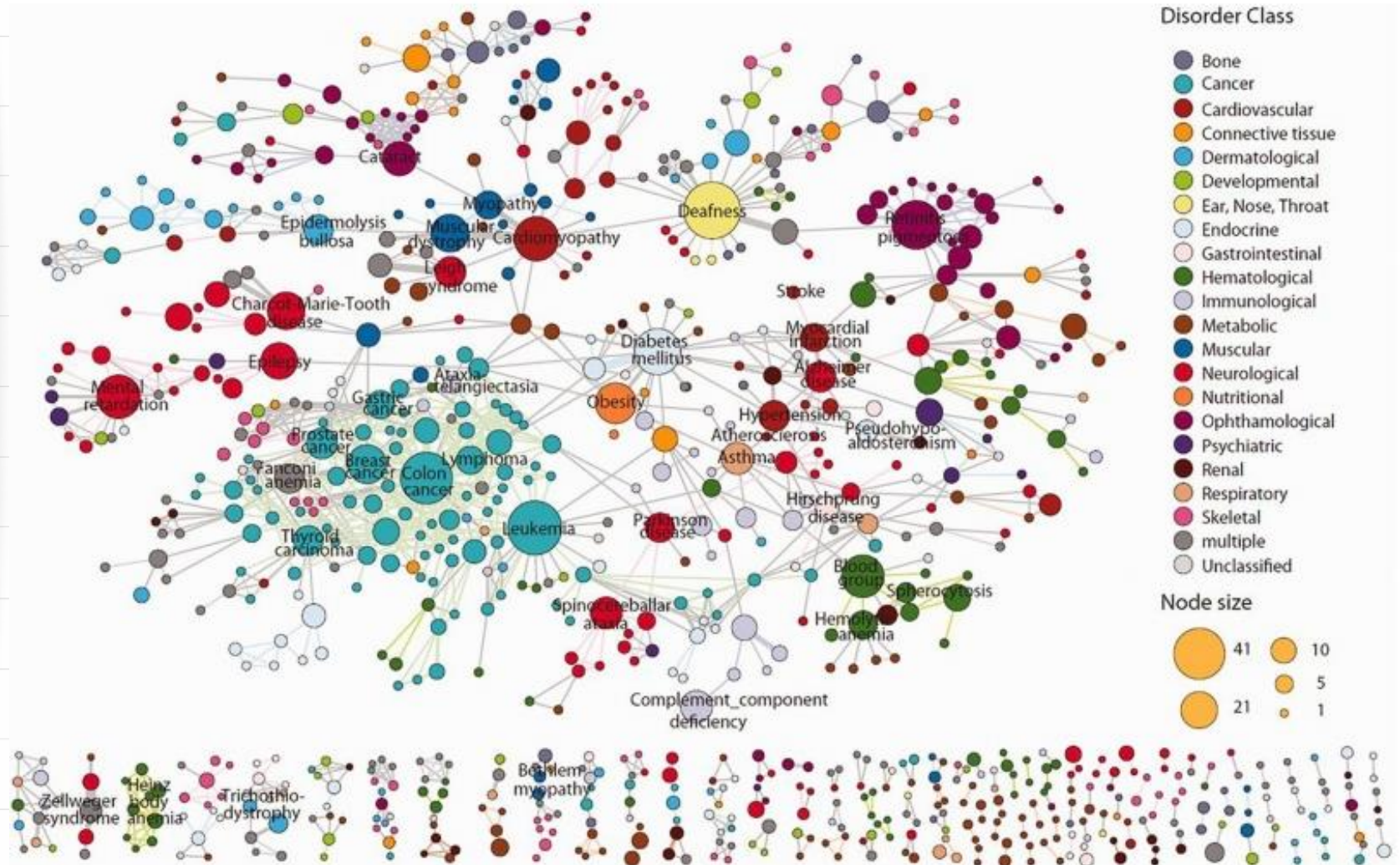
PHENOME

GENOME

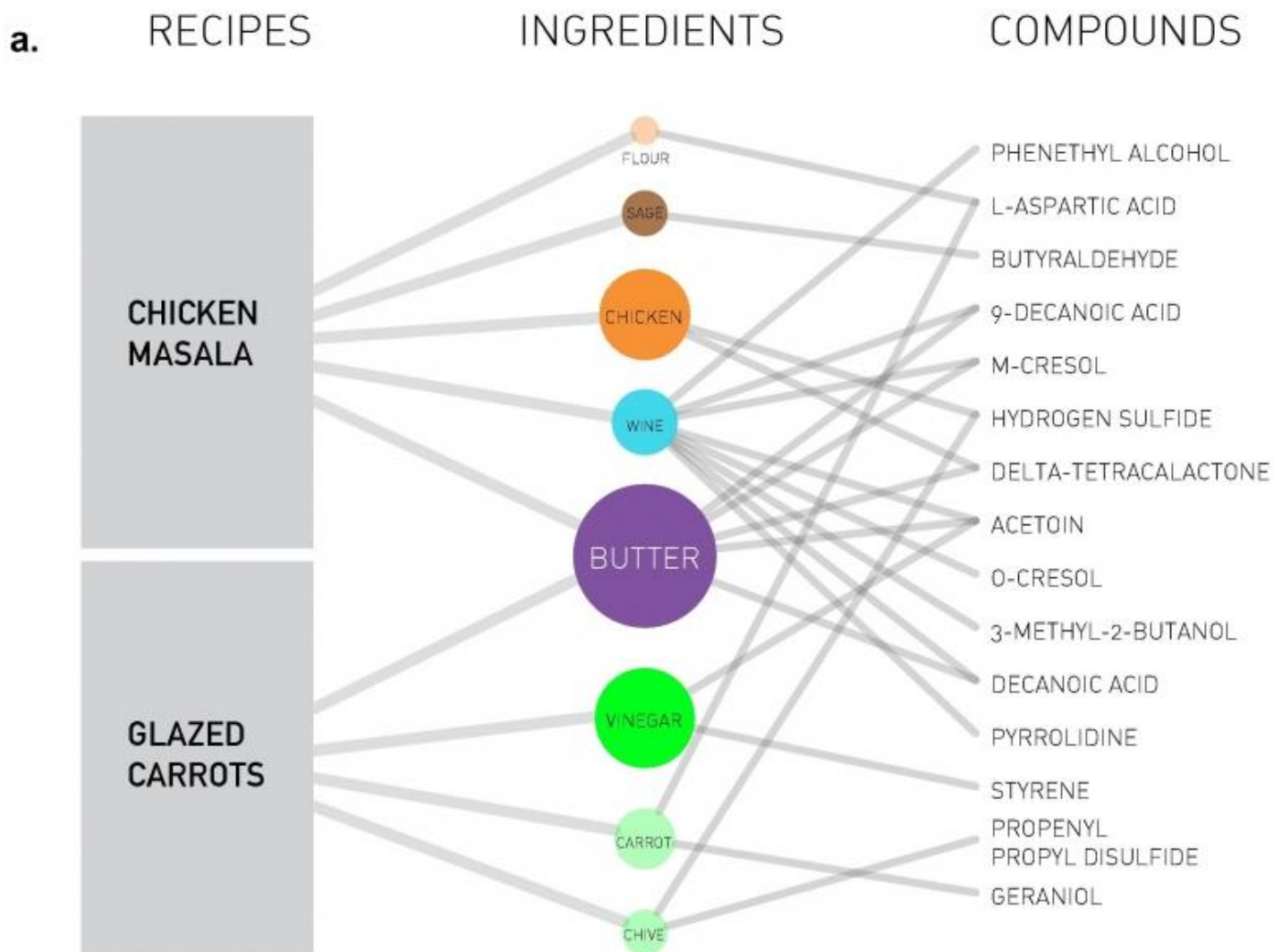


Sieć Chorób

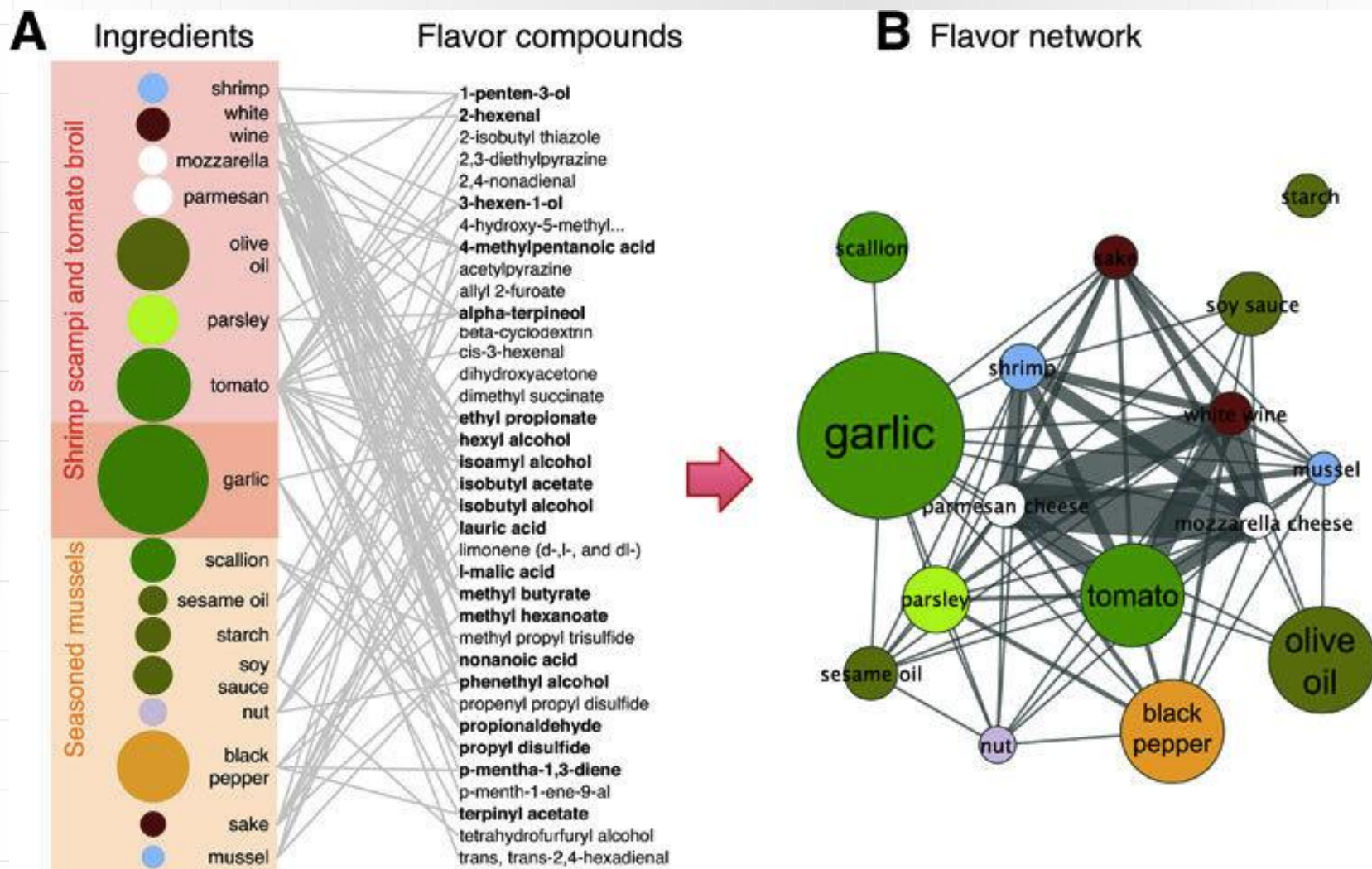
Sieć chorób



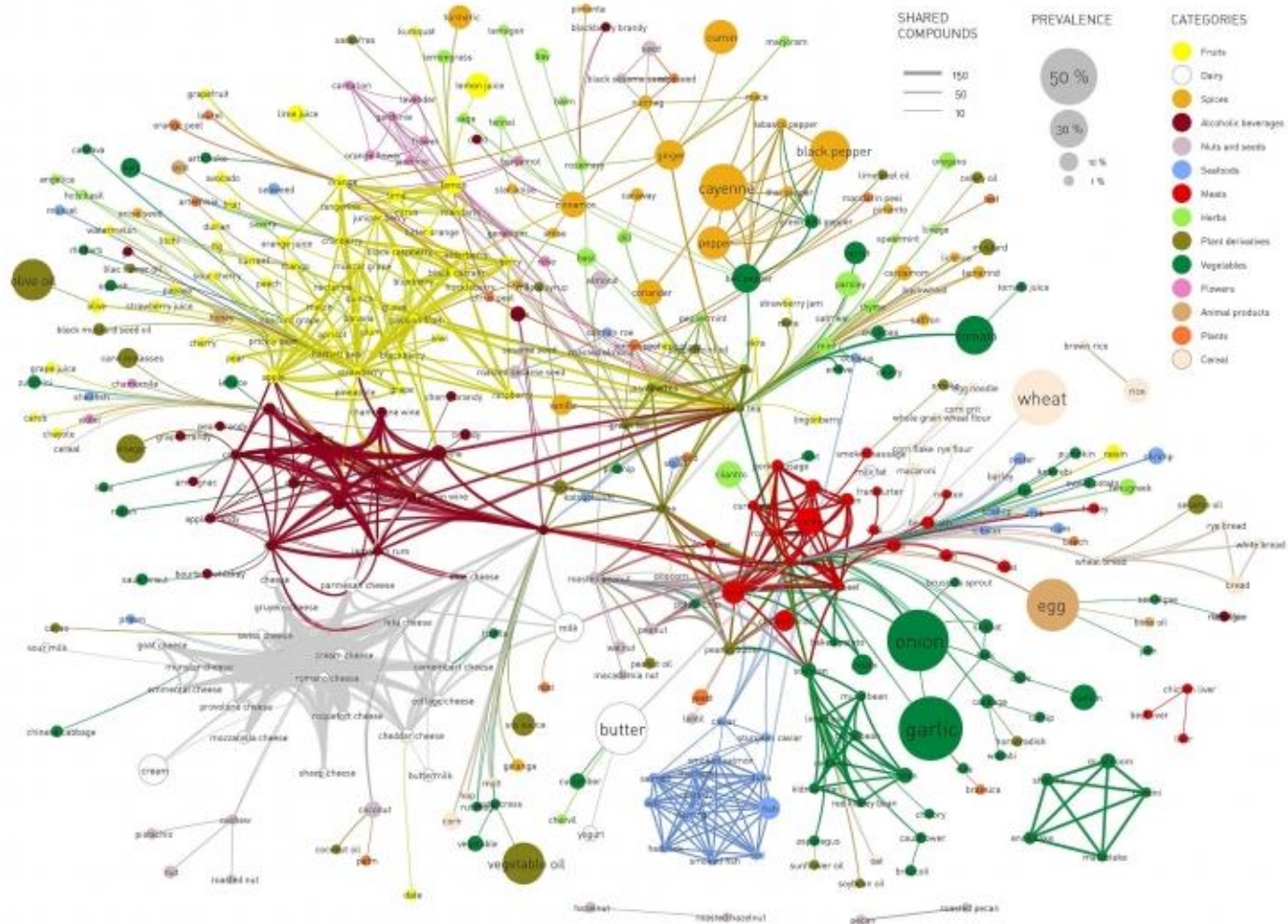
Sieć trójdzielna/Tripartite Network



Sieć trójdzielna



Sieć trójdzielna





Politechnika
Wrocławska

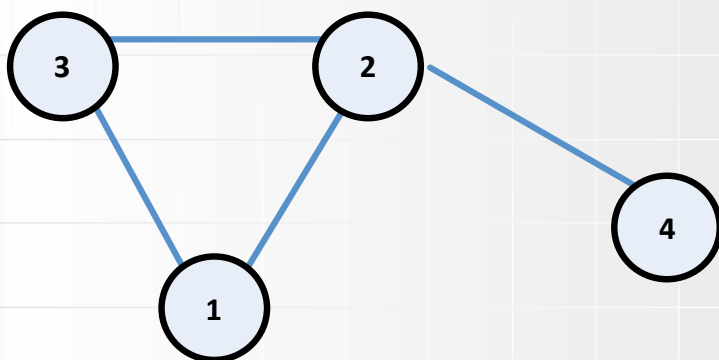
Sieci ważone



HR EXCELLENCE IN RESEARCH

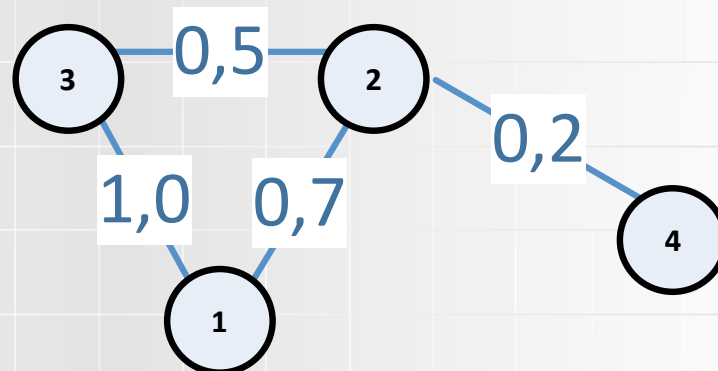
Sieci ważone i nieważone

Nieważona



0	1	1	0
1	0	1	1
1	1	0	0
0	1	0	0

Ważona



0	0,7	1,0	0
0,7	0	0,5	0,2
1,0	0,5	0	0
0	0,2	0	0

$$A_{ij} = w_{ij}$$

- Liczba m³ przesyłana wodociagiem
- Przesłane kWh
- Wartość wymiany handlowej pomiędzy krajami
- Liczba lotów pomiędzy lotniskami
- Stopień zapełnienia autobusów na danej trasie



Politechnika
Wrocławska

Ścieżki w grafie

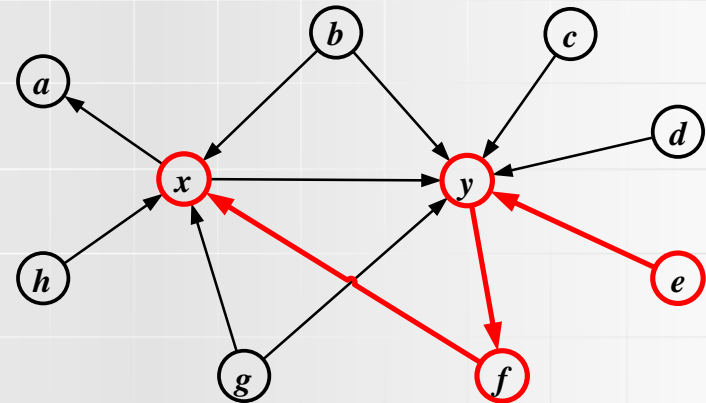
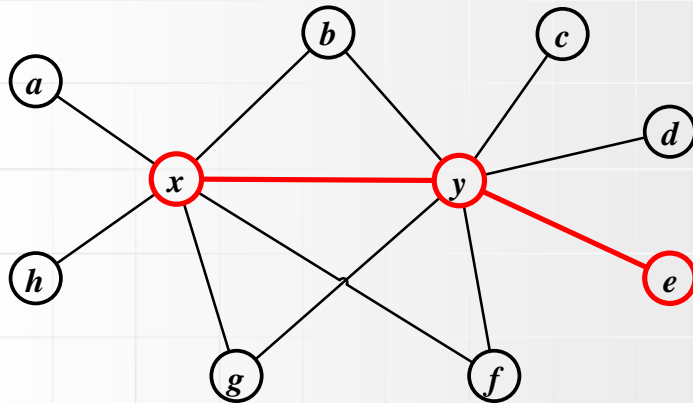


HR EXCELLENCE IN RESEARCH

Ścieżka (Path)

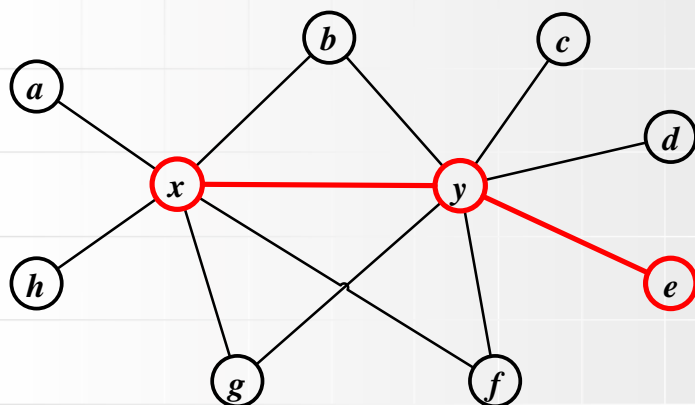
- **Ścieżka** to sekwencja węzłów, w której każdy węzeł sąsiaduje z następnym
- P_{i_0, i_n} o długości n między węzłami i_0 i i_n jest uporządkowanym zbiorem $n+1$ węzłów i n krawędzi

$$P_n = \{i_0, i_1, i_2, \dots, i_n\} \quad P_n = \{(i_0, i_1), (i_1, i_2), (i_2, i_3), \dots, (i_{n-1}, i_n)\}$$

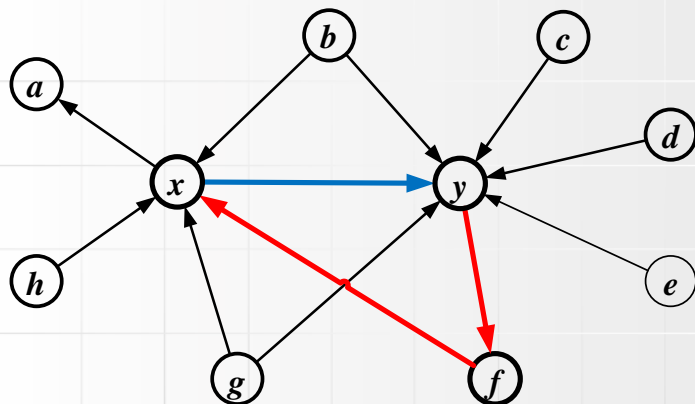


- W grafie skierowanym ścieżka może podążać tylko zgodnie ze zwrotem krawędzi.

Dystans (Distance)



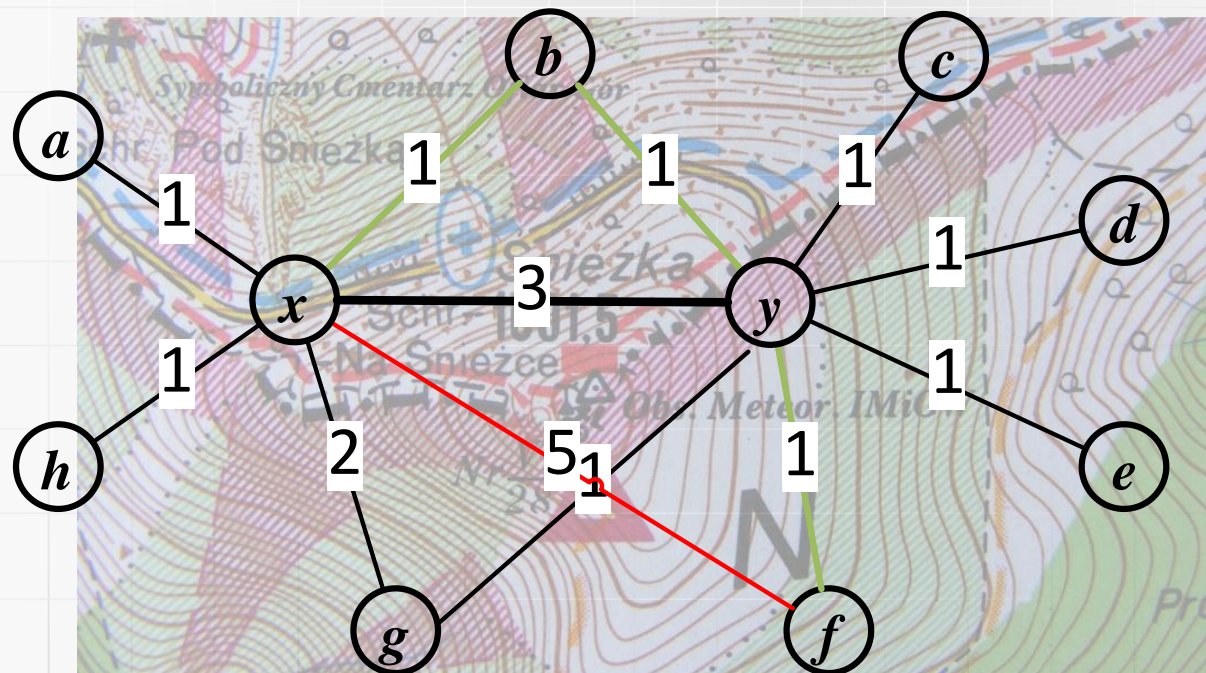
- Odległość (najkrótsza ścieżka, shortest path) między dwoma węzłami jest to liczba krawędzi wzdłuż najkrótszej ścieżki pomiędzy nimi.



- Jeżeli nie istnieje ścieżka pomiędzy dwoma węzłami, odległość jest nieskończona
- Nieskierowane $d(x, e) = d(e, x)$
- Skierowane $d(x, y)$ nie musi być takie samo jak $d(y, x)$

Dystans (Distance)

- W grafach ważonych odległość jest sumą wag krawędzi.
- W poniższym grafie najkrótszą ścieżką od wężła X do F jest ścieżka XBYF, a nie XF czy XYF

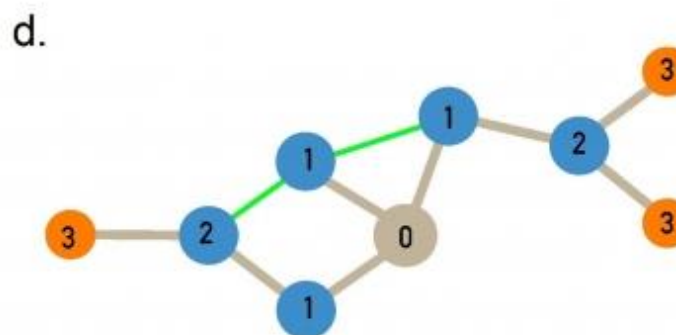
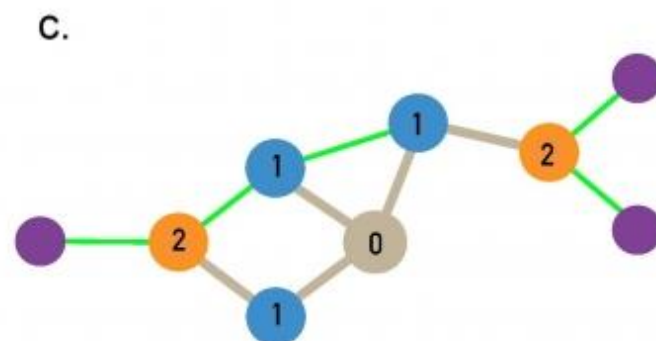
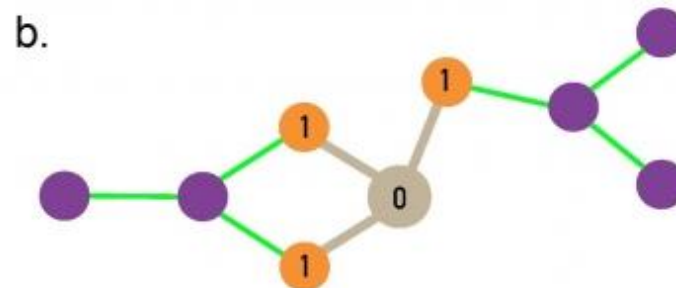
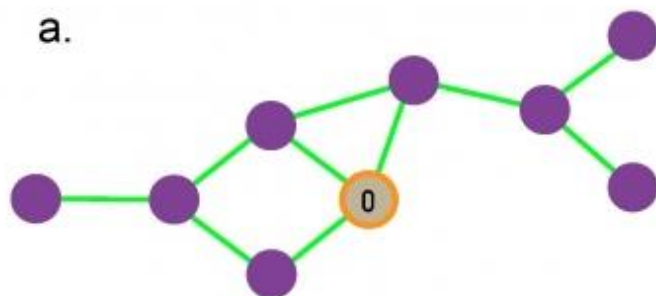


Przeszukiwanie wszerz (BFS - Breadth First Search)

1. Rozpocznij od wierzchołka i , przypisz mu etykietę $m=0$
2. Wybierz wierzchołki bezpośrednio połączone z i . Przypisz im etykietę $m=1$ i umieść w kolejce.
3. Weź pierwszy wierzchołek z kolejki (wierzchołek k o etykiecie m). Wybierz wierzchołki bez etykiety bezpośrednio połączone z k . Przypisz im etykietę $m+1$ i umieść w kolejce.
4. Powtórz krok 3 aż poszukiwany wierzchołek j otrzyma etykietę lub nie ma już wierzchołków w kolejce.

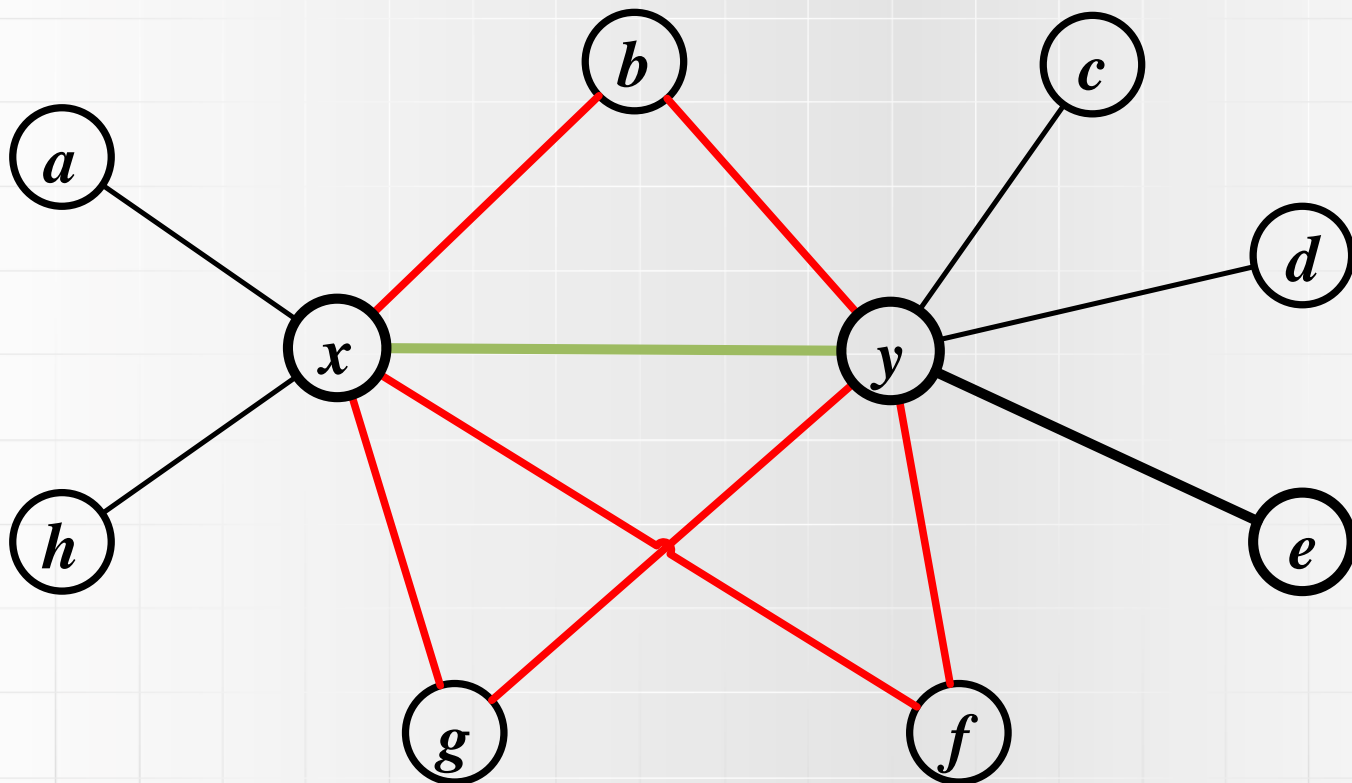
Odległość pomiędzy i oraz j jest równa etykiecie wierzchołka j . Jeżeli j nie ma etykiety wtedy $d(i,j) = \infty$.

Przeszukiwanie wszerz (BFS - Breadth First Search)

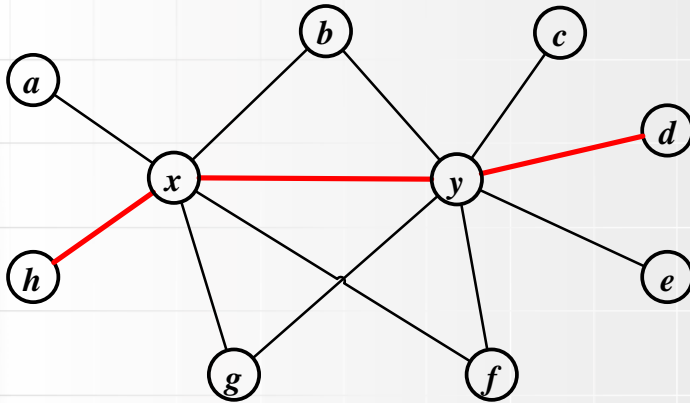


Ścieżki podsumowanie

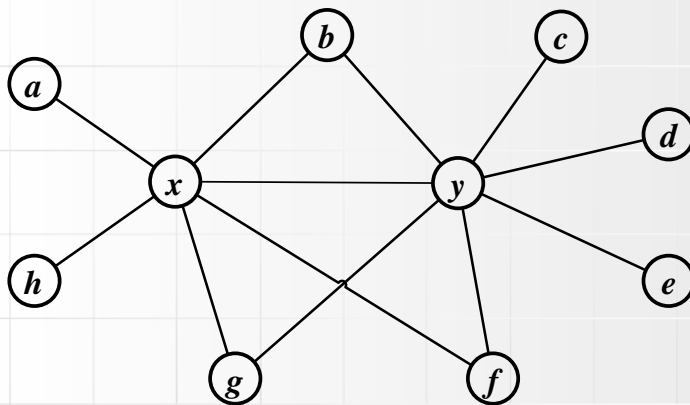
- Najkrótsza ścieżka (Shortest path) – ścieżka pomiędzy dwoma wierzchołkami o najmniejszym dystansie.



Ścieżki podsumowanie

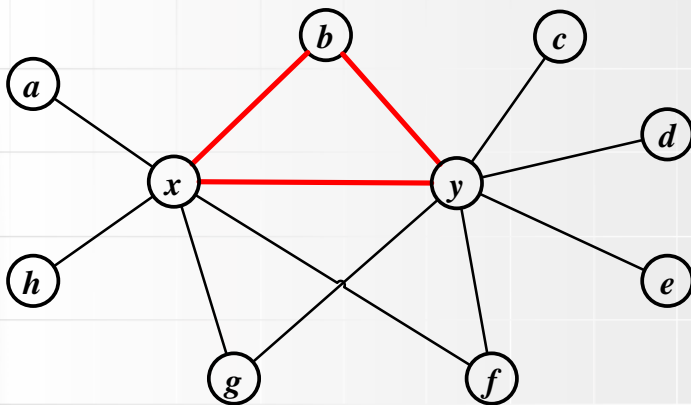


- **Średnica (Diameter)** – najdłuższa najkrótsza ścieżka w grafie. Gdy graf nie jest połączony to średnica = ∞ , lub średnica największego komponentu
- **Średnia długość ścieżki (Average Path Length)** – średnia długość najkrótszych ścieżek pomiędzy wszystkimi parami wierzchołków w grafie.

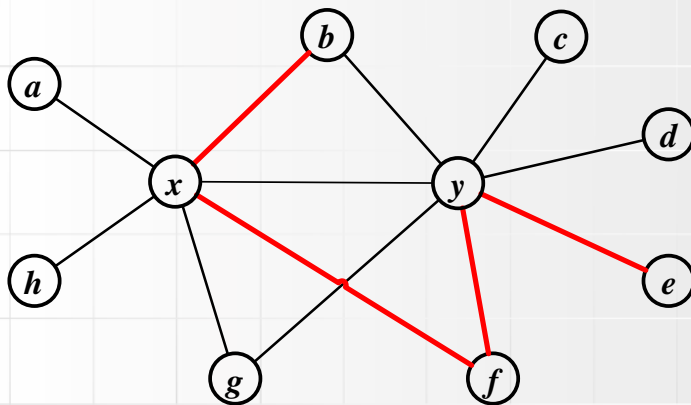


90 ścieżek – APL = 1.867

Ścieżki podsumowanie



- Cykl (Cycle) – ścieżka która zaczyna się i kończy w tym samym wierzchołku



- Samounikająca ścieżka (Self-avoiding Path) – ścieżka która się nie krzyżuje sama ze sobą

Ścieżki podsumowanie

- Ścieżka **Eulera** (**Eulerian Path**) – ścieżka która przechodzi przez wszystkie **krawędzie** (**Edges**) w grafie dokładnie jeden raz
- Ścieżka **Hamiltona** (**Hamiltonian path**) – ścieżka która przechodzi przez wszystkie **wierzchołki** dokładnie jeden raz



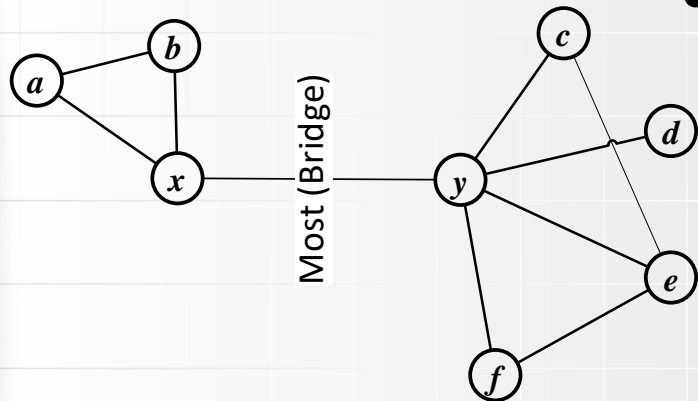
Politechnika
Wrocławska

Połączenie (Connectedness)

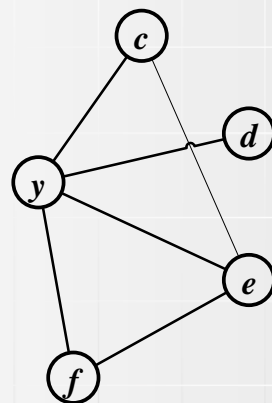
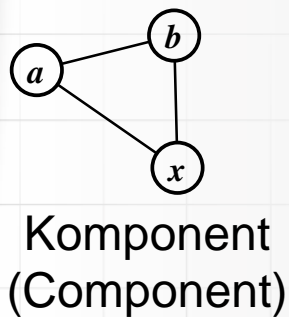


HR EXCELLENCE IN RESEARCH

Graf połączony



- Graf połączony (nieskierowany) występuje wtedy gdy można wyznaczyć ścieżkę pomiędzy dwoma dowolnymi wierzchołkami.

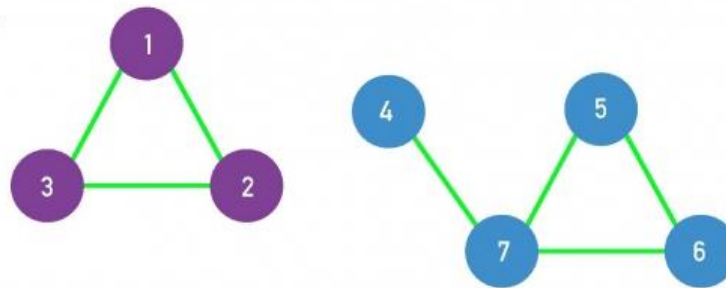


- Graf niepołączony (nieskierowany) Występuje wtedy gdy występuje co najmniej jedna para wierzchołków między którymi nie można wytyczyć ścieżki.

Graf połączony

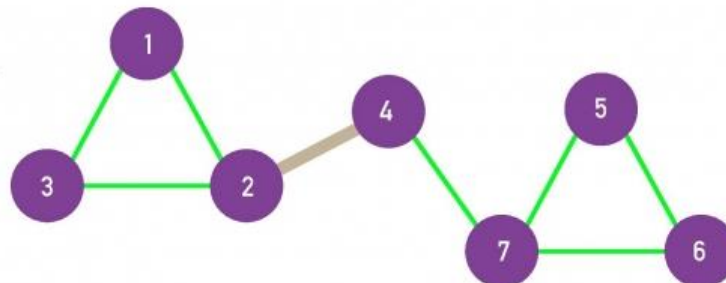
- Macierz sąsiedztwa sieci składającej się z kilku elementów może być zapisana w postaci blokowo-przekątnej, tak że niezerowe elementy są ograniczone do kwadratów, a wszystkie inne elementy są zerowe

a.



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

b.



$$\begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 \end{pmatrix}$$

Graf połączony

- Silnie połączony **graf skierowany**: każdy węzeł ma ścieżkę od każdego wężła.
- Słabo połączony **graf skierowany**: jest połączony, jeśli pominiemy kierunki krawędzi (zrobimy projekcję na graf nieskierowany).

Znajdowanie podłączonych komponentów

1. Rozpocznij od losowego wierzchołka j i wykonaj dla niego algorytm BFS. Wszystkim wierzchołom do których dotarłeś nadaj etykietę $m = 1$.
2. Jeżeli liczba wierzchołków z etykietą wynosi N to graf jest połączony natomiast jeżeli jest mniejsza niż N to graf składa się z kilku komponentów. By je zidentyfikować przejdź do kroku 3.
3. Rozpocznij od losowego wierzchołka k bez etykiety i wykonaj dla niego algorytm BFS. Wierzchołom do których dotarłeś nadaj etykietę $m = m + 1$.
4. Jeżeli liczba wierzchołków z etykietą wynosi N to liczba komponentów wynosi m . Jeżeli liczba wierzchołków z etykietą jest mniejsza od N wróć do kroku 3.



Politechnika
Wrocławska

Współczynnik grupowania

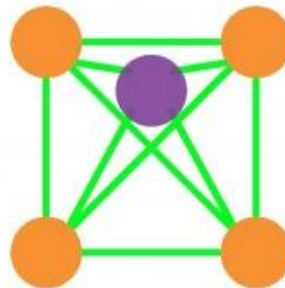


HR EXCELLENCE IN RESEARCH

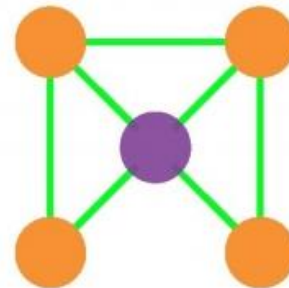
Lokalny współczynnik grupowania

- Współczynnik grupowania (Clustering coefficient) C_i mówi nam jak dobrze połączone jest sąsiedztwo wierzchołka i
- Jeżeli stopień wężła k_i wynosi 0 lub 1 to współczynnik grupowania wynosi 0

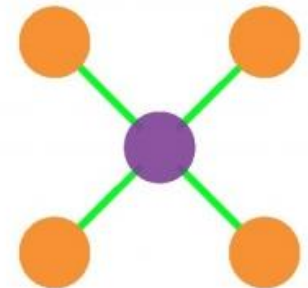
$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



$$C_i = 1$$



$$C_i = 1/2$$



$$C_i = 0$$

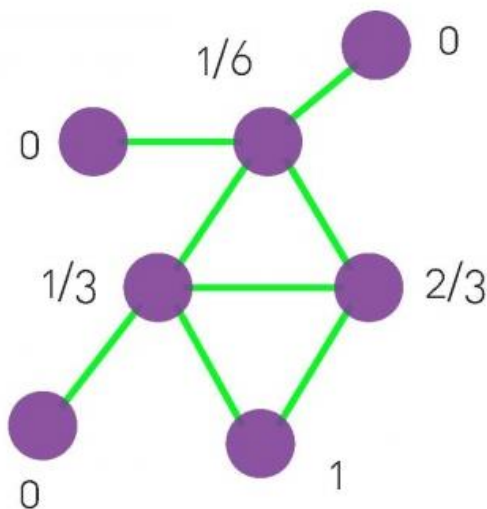
Współczynnik grupowania

- Średni współczynnik grupowania (average clustering coefficient)

$$\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$$

- Globalny współczynnik grupowania (global clustering clustering)

$$C_{\Delta} = \frac{3 \times \text{Liczba trójkątów}}{\text{Liczba połączonych trójek}}$$



$$\langle C \rangle = \frac{13}{42} \approx 0.310$$

$$C_{\Delta} = \frac{3}{8} = 0.375$$



Politechnika
Wrocławska

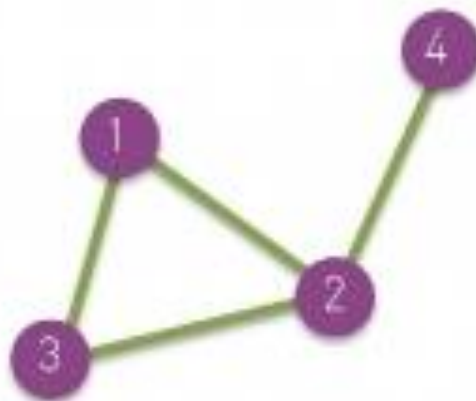
Grafy podsumowanie



HR EXCELLENCE IN RESEARCH

Nieskierowane

a. Undirected



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

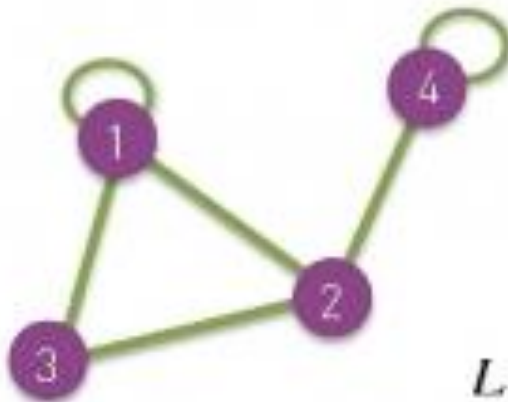
$$A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

Nieskierowane z pętlami

b. Self-loops



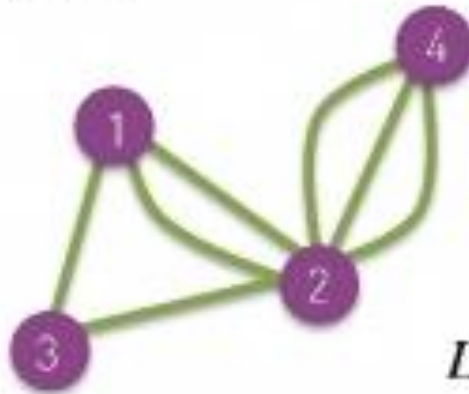
$$A_{ij} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

$$L = \frac{1}{2} \sum_{i,j=1, i \neq j}^N A_{ij} + \sum_{i=1}^N A_{ii}$$

$\exists i, A_{ii} \neq 0$ $A_{ij} = A_{ji}$?

Multigrafy

c. Multigraph (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 2 & 1 & 0 \\ 2 & 0 & 1 & 3 \\ 1 & 1 & 0 & 0 \\ 0 & 3 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

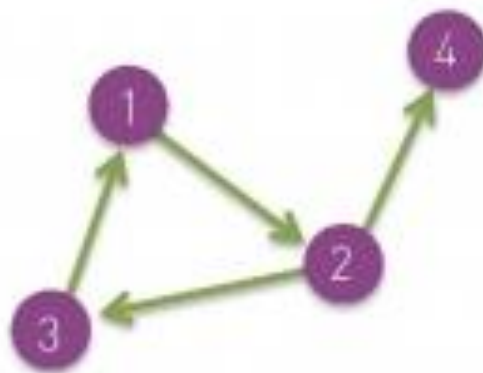
$$A_{ij} = A_{ji}$$

$$L = \frac{1}{2} \sum_{i,j=1}^N A_{ij}$$

$$\langle k \rangle = \frac{2L}{N}$$

Skierowane

d. Directed



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}$$

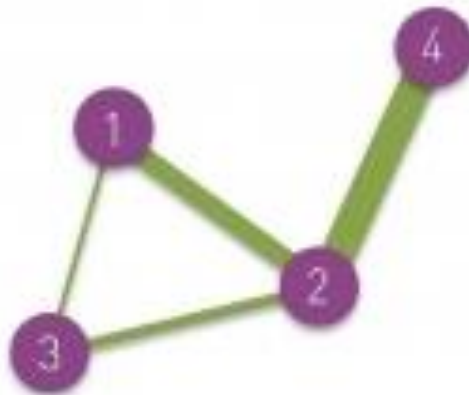
$$A_{ij} \neq A_{ji}$$

$$L = \sum_{i,j=1}^N A_{ij}$$

$$\langle k \rangle = \frac{L}{N}$$

Ważone

e. Weighted
(undirected)



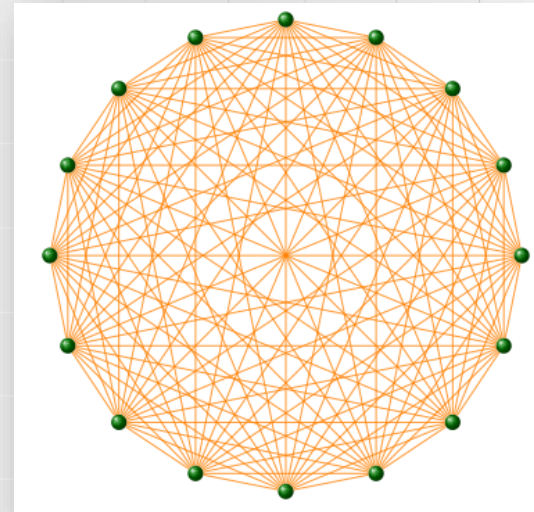
$$A_{ij} = \begin{pmatrix} 0 & 2 & 0.5 & 0 \\ 2 & 0 & 1 & 4 \\ 0.5 & 1 & 0 & 0 \\ 0 & 4 & 0 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

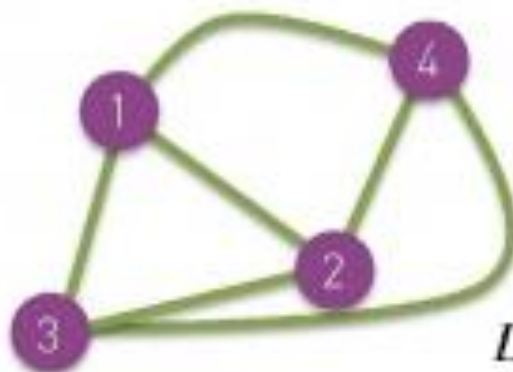
$$A_{ij} = A_{ji}$$

$$\langle k \rangle = \frac{2L}{N}$$

Pełne



f. Complete Graph (undirected)



$$A_{ij} = \begin{pmatrix} 0 & 1 & 1 & 1 \\ 1 & 0 & 1 & 1 \\ 1 & 1 & 0 & 1 \\ 1 & 1 & 1 & 0 \end{pmatrix}$$

$$A_{ii} = 0$$

$$A_{i \neq j} = 1$$

$$L = L_{\max} = \frac{N(N-1)}{2}$$

$$\langle k \rangle = N - 1$$

Przykłady

- WWW
 - skierowane multigrafy z pętlami
- Interakcje pomiędzy białkami
 - Nieskierowane, nieważone z pętlami
- Sieć współautorstwa
 - Nieskierowane, ważne multigrafy
- Połączenia telefoniczne
 - Skierowane, ważne, multigrafy
- „Przyjaźń” na Facebooku
 - Nieskierowane, nieważone



Politechnika
Wrocławska

Główne miary centralności



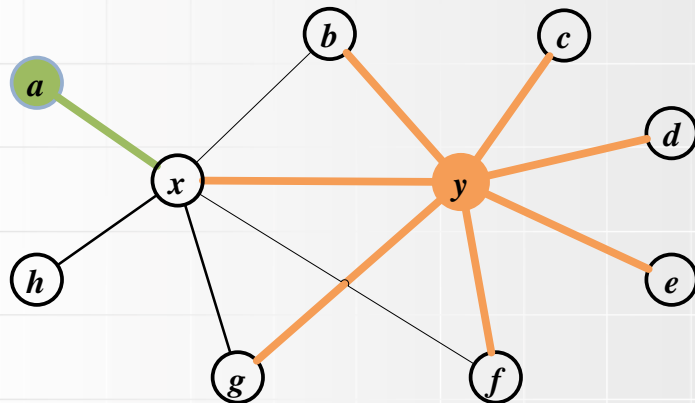
HR EXCELLENCE IN RESEARCH

Centralność

- **Centralność (centrality) identyfikuje najważniejsze z jakiegoś punktu widzenia wierzchołki w grafie.**
- Stopień wężła (degree)
- Przewodnictwo (betweenness)
- Bliskość (closeness)
- PageRank
- eigenvector (prestige score)
- Katz
- K-shell
- VoteRank
-

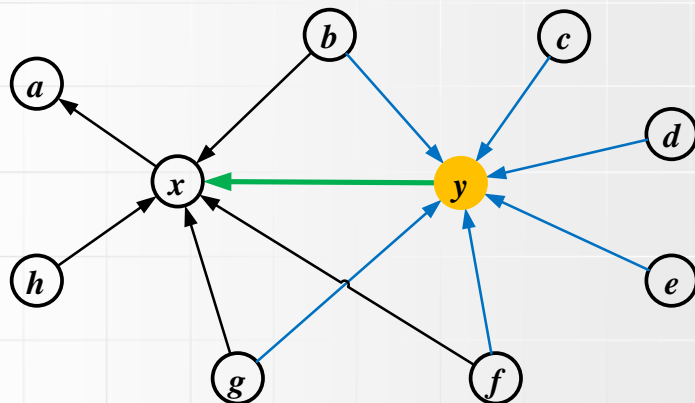
Stopień Węzła k_i

Nieskierowana



$$k_a = 1, k_y = 7$$

Skierowana



$$k_y^{in} = 6 \quad k_y^{out} = 1 \quad k_y = 7$$

Stopień węzła: liczba krawędzi połączonych z węzłem.

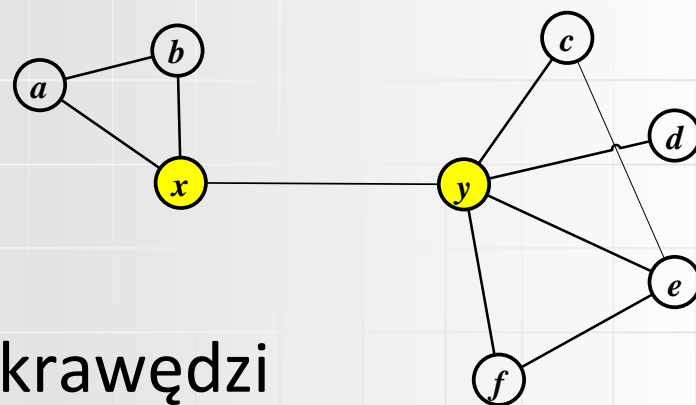
Przewodnictwo

- **Przewodnictwo (Betweenness centrality)** określa pozycję wężła na podstawie liczby najkrótszych ścieżek przechodzących przez ten węzeł w stosunku do liczby wszystkich najkrótszych ścieżek.

$$C_B(v) = \sum_{s \neq v \neq t \in V} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

liczba tych ścieżek, które przechodzą przez v

całkowita liczba najkrótszych ścieżek od wężła s do wężła t



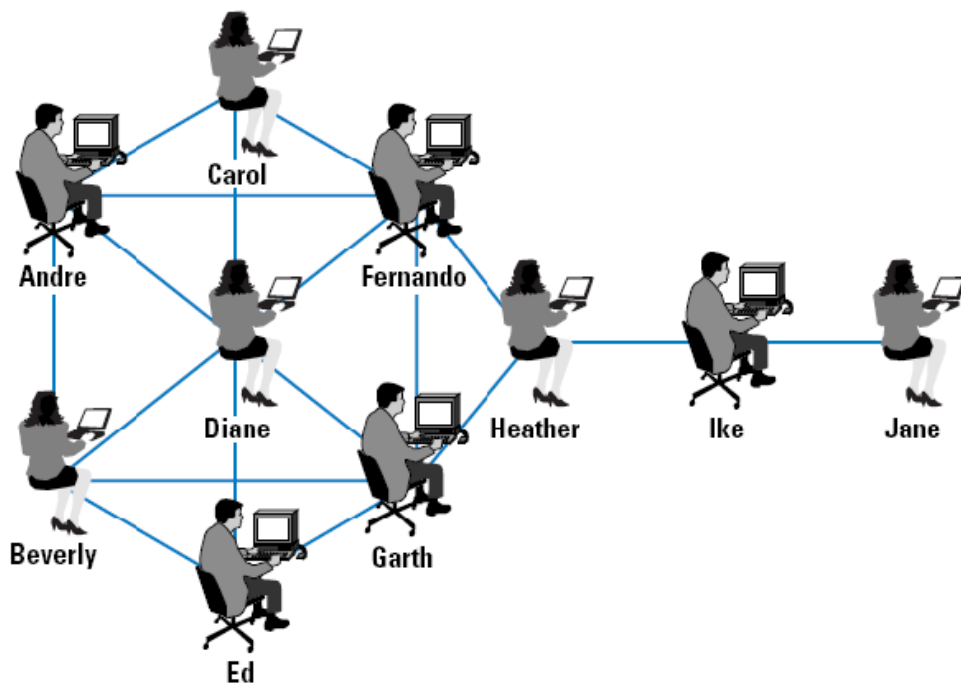
- Także może być policzona dla krawędzi

Bliskość

- W połączonym grafie bliskość (closeness centrality) to średnia długość najkrótszej ścieżki między węzłem a wszystkimi innymi węzłami w grafie.
- Im bardziej centralny węzeł, tym znajduje się „bliżej” wszystkich innych węzłów.

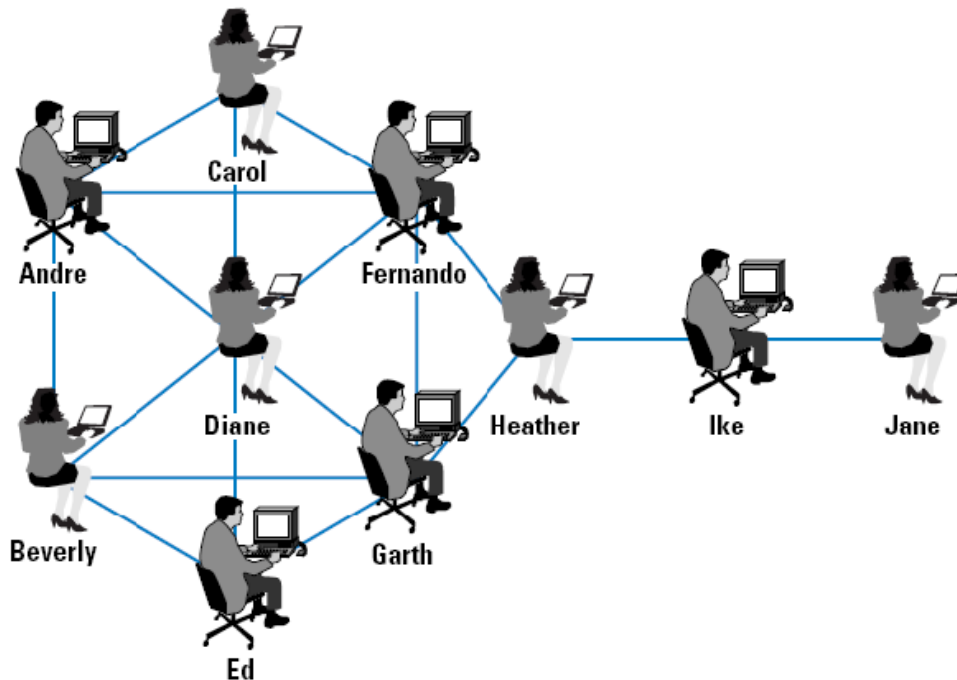
$$C(x) = \frac{1}{\sum_y d(y, x)}$$

Przykład



	C_D	C_C	C_B
Diane			
Fernando			
Garth			
Andre			
Beverly			
Carol			
Ed			
Heather			
Ike			
Jane			

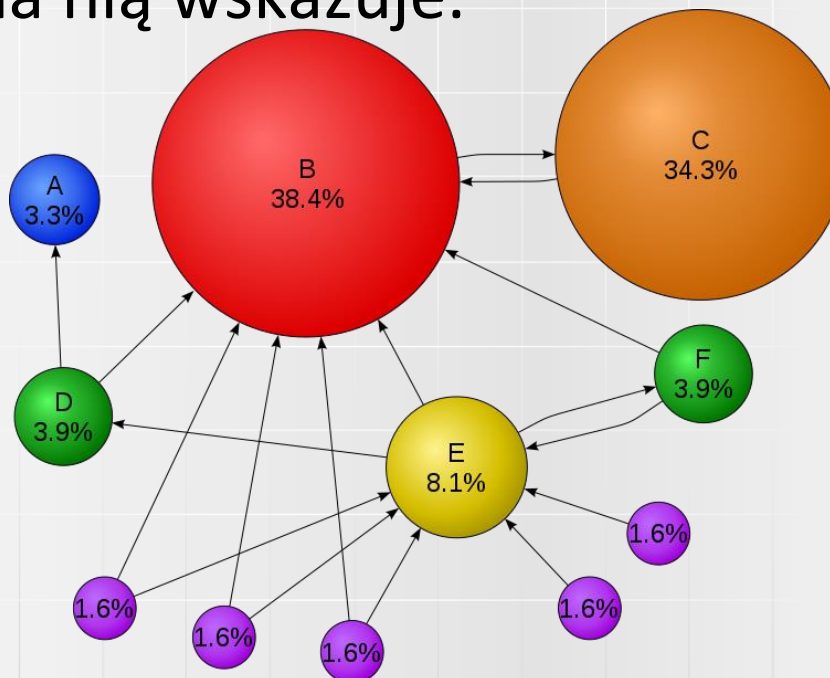
Przykład



	C_D	C_C	C_B
Diane	0.666	0.600	0.102
Fernando	0.556	0.643	0.231
Garth	0.556	0.643	0.231
Andre	0.444	0.529	0.023
Beverly	0.444	0.529	0.023
Carol	0.333	0.500	0.000
Ed	0.333	0.500	0.000
Heather	0.333	0.600	0.389
Ike	0.222	0.429	0.222
Jane	0.111	0.310	0.000

PageRank

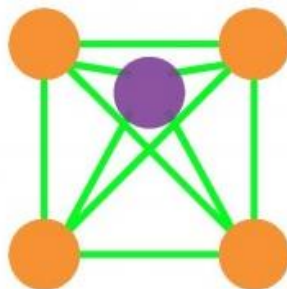
- PageRank polega na liczeniu liczby i jakości linków do strony (węzła), aby określić przybliżoną ocenę ważności witryny (węzła). Podstawowym założeniem jest to, że im strona jest ważniejsza tym więcej odnośników na nią wskazuje.



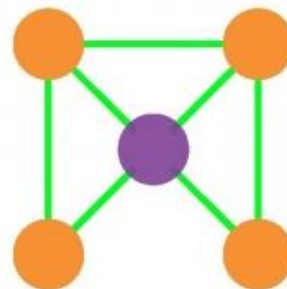
Lokalny współczynnik grupowania

- Współczynnik grupowania (Clustering coefficient) C_i mówi nam jak dobrze połączone jest sąsiedztwo wierzchołka i
- Jeżeli stopień wężła k_i wynosi 0 lub 1 to współczynnik grupowania wynosi 0

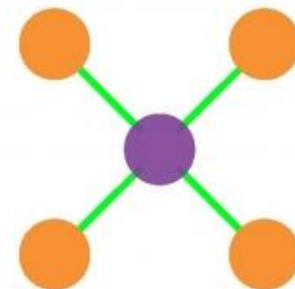
$$C_i = \frac{2e_i}{k_i(k_i - 1)}$$



$$C_i = 1$$



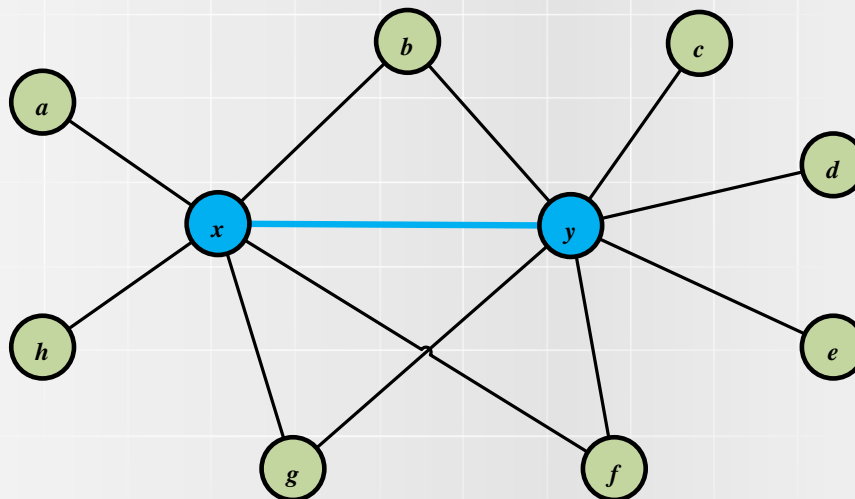
$$C_i = 1/2$$



$$C_i = 0$$

Współczynnik grupowania dla krawędzi

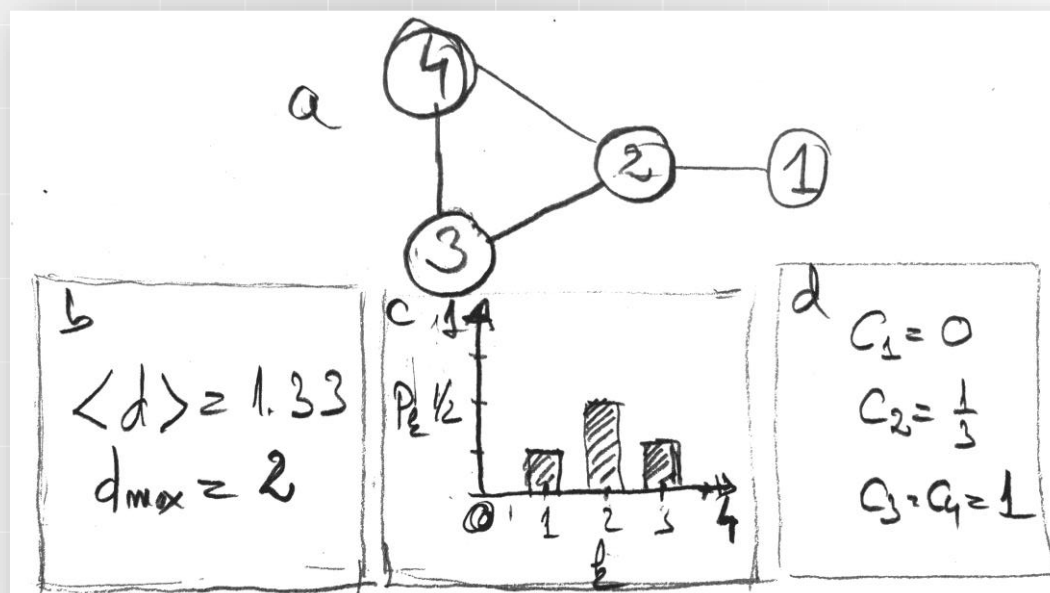
$$CLECC(x, y, \alpha) = \frac{|MN(x, \alpha) \cap MN(y, \alpha)|}{|(MN(x, \alpha) \cup MN(y, \alpha)) / \{x, y\}|}$$



$$\begin{aligned} CLECC(x, y, 1) &= \frac{|MN(x, 1) \cap MN(y, 1)|}{|(MN(x, 1) \cup MN(y, 1)) / \{x, y\}|} = \frac{|\{a, b, y, g, h\} \cap \{b, c, d, e, f, g, x\}|}{|\{a, b, y, g, h\} \cup \{b, c, d, e, f, g, x\}| / \{x, y\}|} = \\ &= \frac{|\{b, f, g\}|}{|\{a, b, c, d, e, f, g, h, x, y\}| / \{x, y\}|} = \frac{3}{|\{a, b, c, d, e, f, g, h\}|} = \frac{3}{8} \end{aligned}$$

Trzej królowie

- Stopień wężła, średni stopień i rozkład stopni wężła
- Długości najkrótszych ścieżek, ich rozkład i średnia długość ścieżki
- Współczynnik grupowania globalny i lokalny i rozkład współczynników grupowania





Politechnika
Wrocławska

Typy sieci a grafy

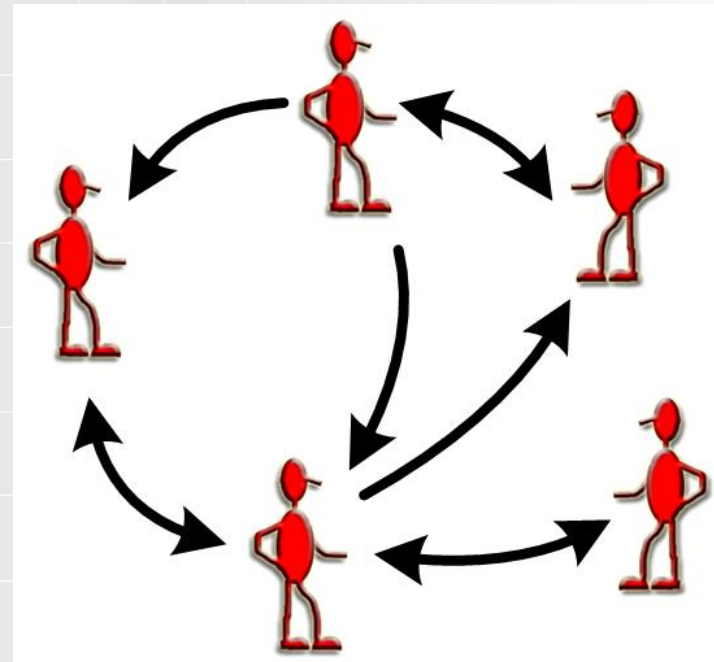


HR EXCELLENCE IN RESEARCH

Sieć prosta

- Jeden typ relacji (jednowarstwowa)

- Uniplex
- Monoplex
- Singleplex
- Singlex
- Simplex
- One layer
- Single layer

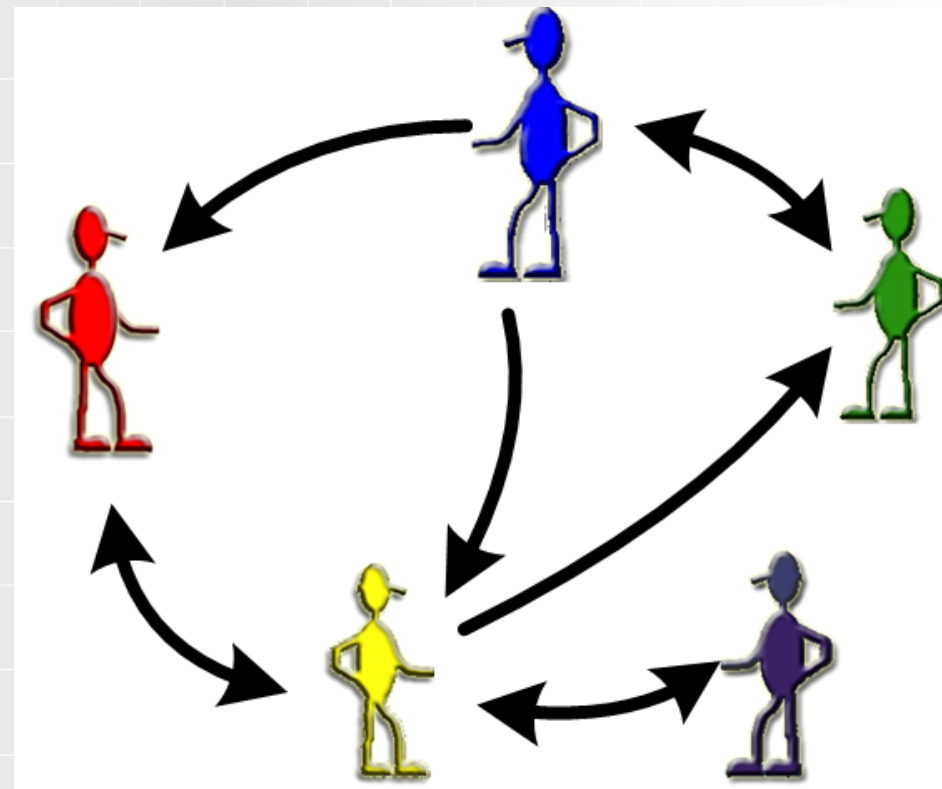


- Jednomodowa (te same wierzchołki)

- Homogenous nodes
- Unimodal

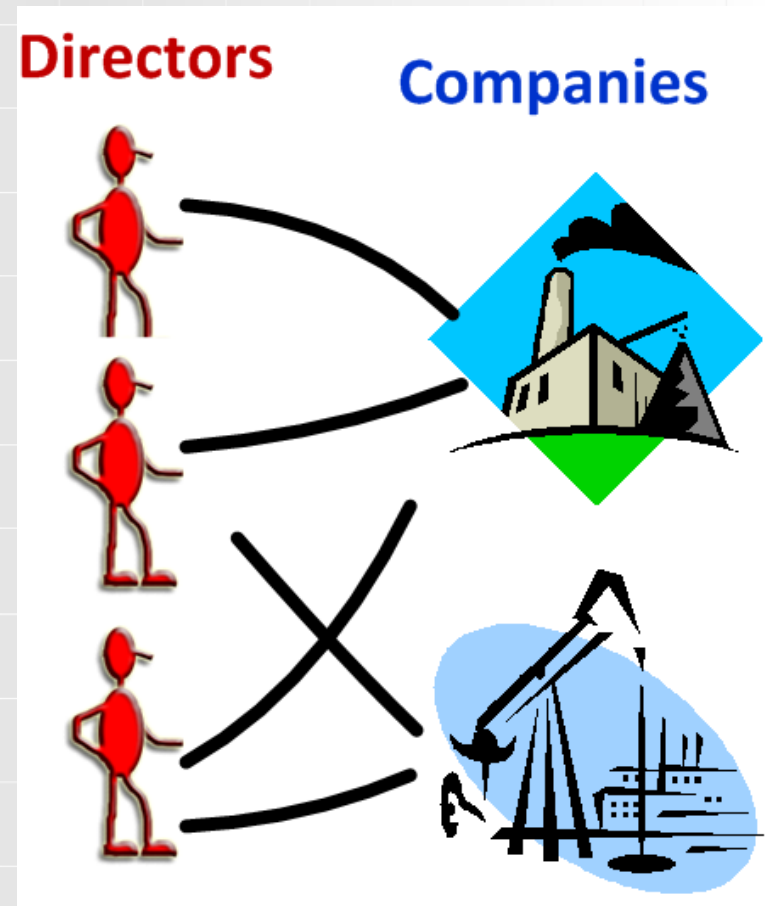
Sieć prosta wielomodowa

- Jeden tryb relacji (uniplex)
- Wiele typów wierzchołków (multimodal)



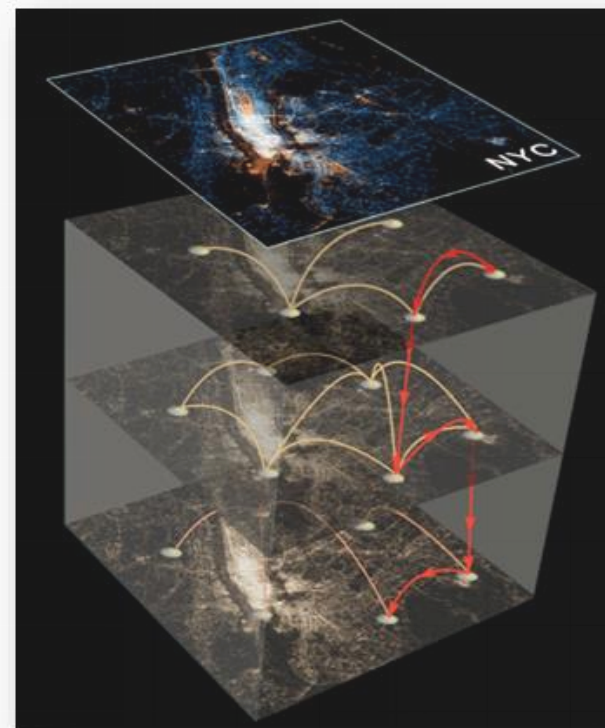
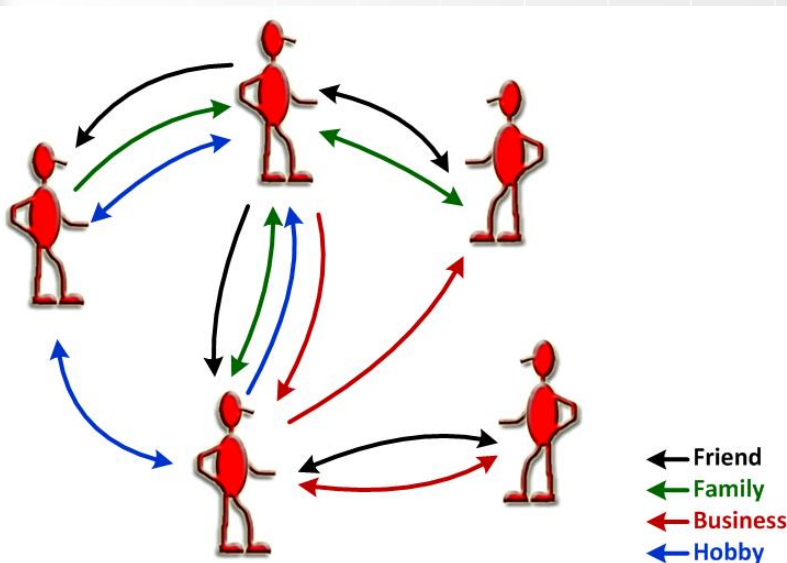
Sieć Xdzielna

- X typów wierzchołków (multimodal)
- Brak połączeń między wierzchołkami tego samego typu



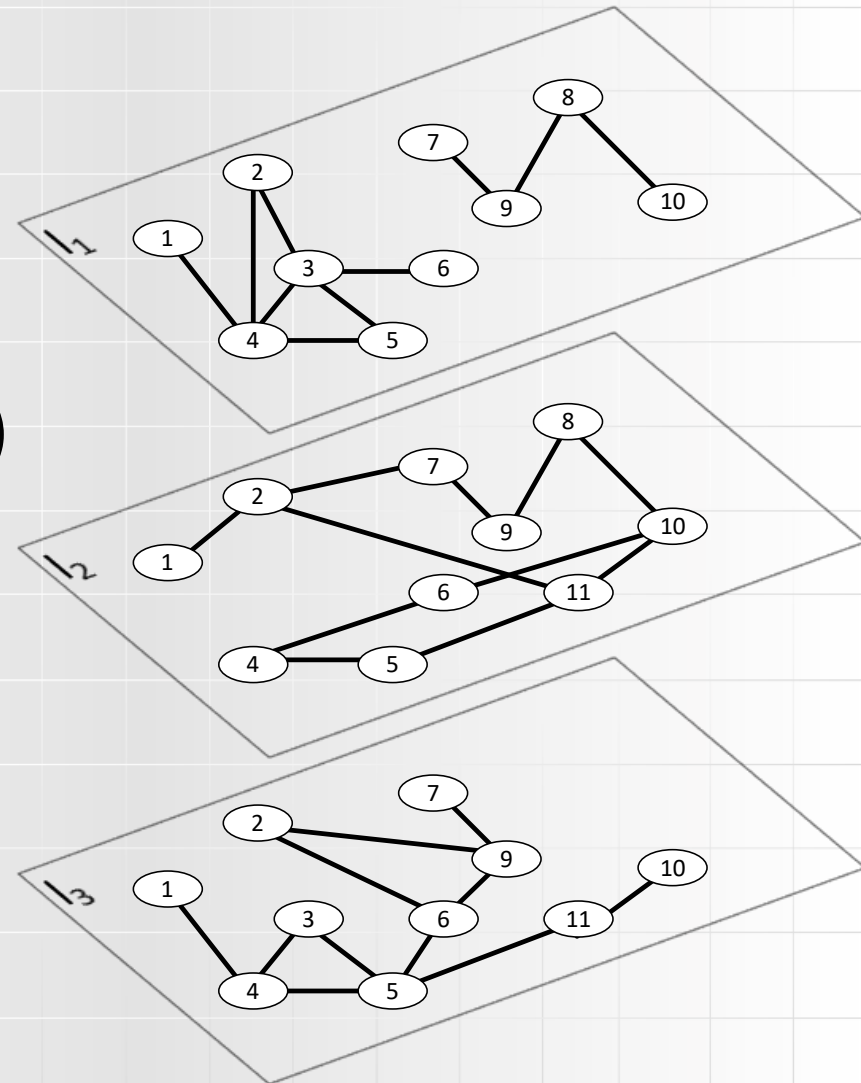
Sieć multipleksowa (Multiplex Network)

- Każda warstwa to sieć prosta o:
 - Dokładnie tym samym zbiorze wierzchołków (node-align multiplex)*
 - Różne zbiory krawędzi
 - np. multigraf



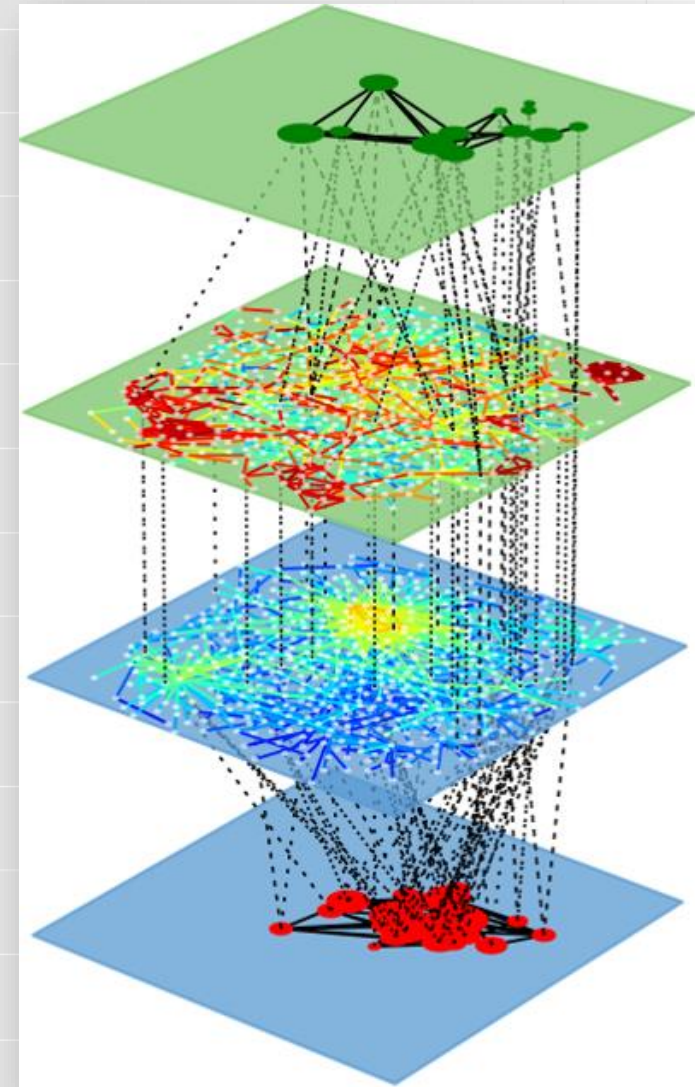
Sieć wielowarstwowa (Multilayer Network)

- Każda warstwa to sieć prosta
- Każda warstwa może mieć inny zbiór wierzchołków i krawędzi (intralayer edges)
- Mogą występować relacje pomiędzy warstwami (interlayer edges)



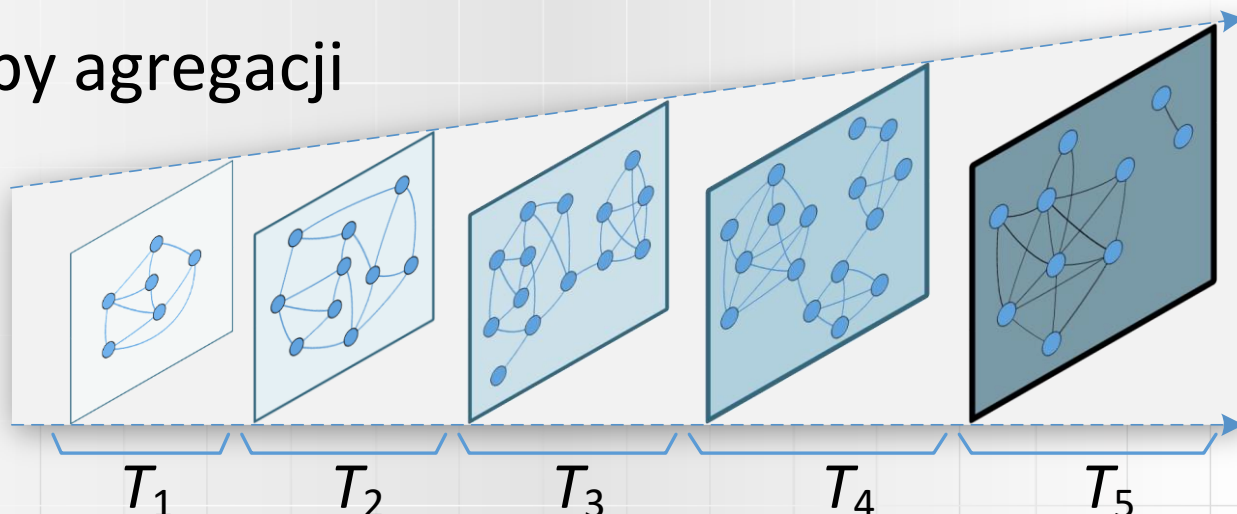
Sieć sieci (Network of networks)

- Każda warstwa to sieć prosta
- Mogą być różne typy wierzchołków na każdej warstwie (multimodal)
- Różne typy relacji w ramach warstwy
- Różne typy relacji pomiędzy warstwami



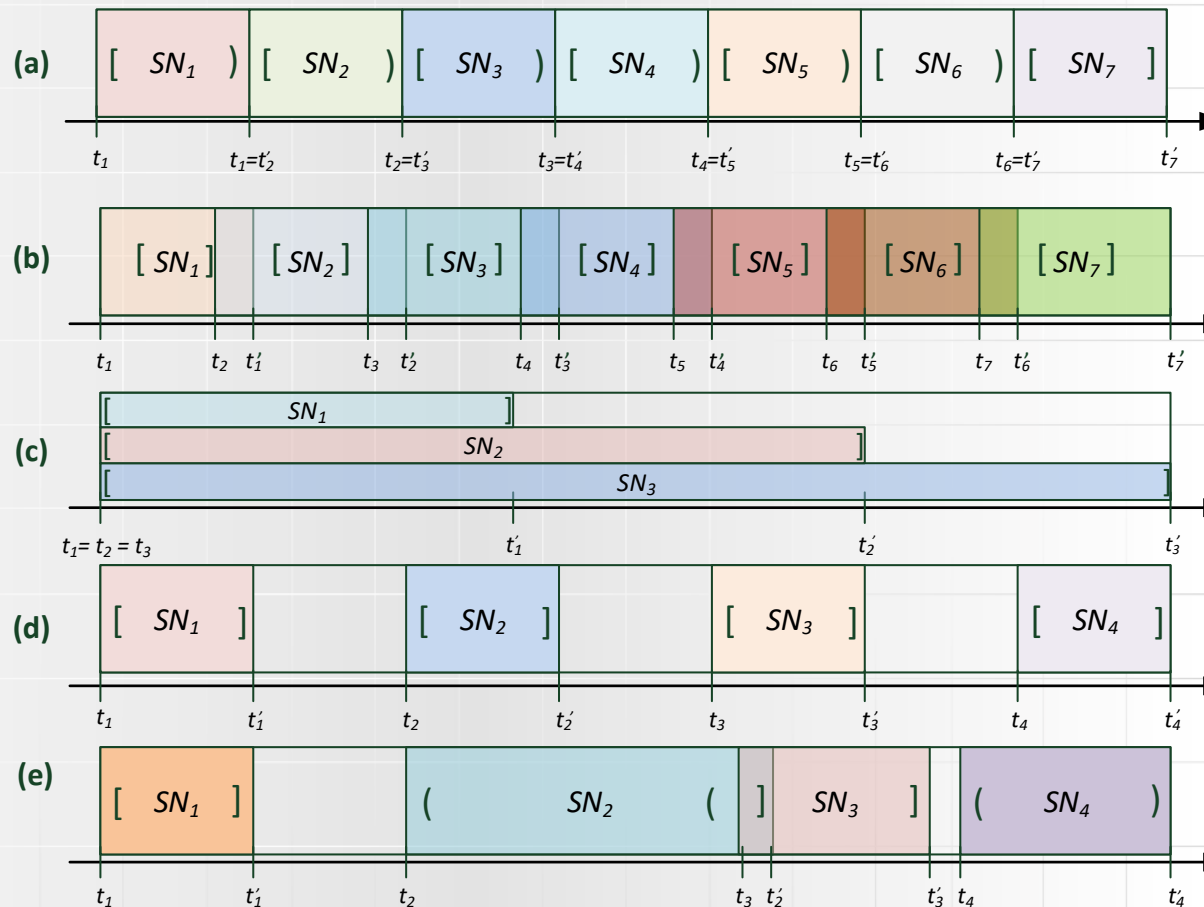
Sieci teporalne (Temporal Network)

- Sieć uwzględniając a czas i uwzględnia ewolucję sieci
- Agregacja do „okien czasowych” (time frames, time windows, snapshots)
- Każde okno to sieć agregująca relacje z zadanego okresu
- Mogą być postrzegana jako sieć wielowarstwowa
- Różne sposoby agregacji



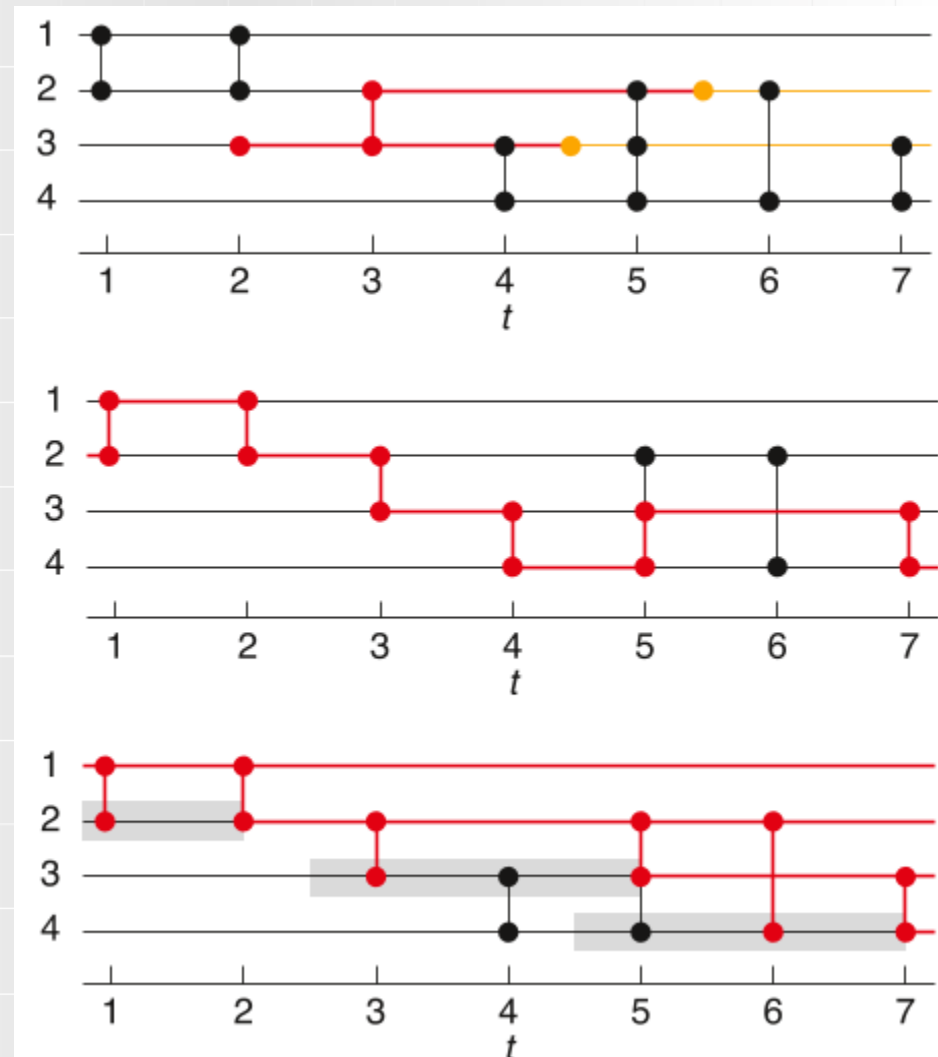
Sieci temporalne (Temporal Network)

- a) Okna sąsiadujące
- b) Okna nachodzące
- c) Okna rozszerzające się
- d) Okna sąsiadujące z przerwą
- e) Mieszane



Sieci temporalne (Temporal Network)

- Sieć uwzględniając a czas i uwzględnia ewolucję sieci.
- Przetwarzane jako strumień danych bez agregacji do „klasycznej” sieci



Następne laboratorium

- Jeżeli w trakcie laboratorium chcecie używać własnego sprzętu proszę upewnić się że macie:
 - a) Działające środowisko do programowania w Pythonie wraz z biblioteką NetworkX i zależnymi (<https://networkx.org/>)
 - b) Działające środowisko do programowania w R wraz z biblioteką igraph i zależnymi (<https://igraph.org/>)
- Można także wcześniej pobrać ze strony <http://konect.cc/networks/>
 - Sieć http://konect.cc/networks/radoslaw_email/

Pytania

