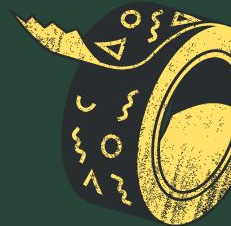




# Wstęp do Eksploracji Danych

Politechnika Warszawska

Anna Kozak



**Wizualizacja grafów - jak nie zgubić się  
w gąszczu wierzchołków i krawędzi**

# Grafy

- graph  $G = (V, E)$
- vertices  $V = \{v_1, v_2, \dots, v_n\}$
- edge  $E = \{e_1, e_2, \dots, e_m\}$

# Reprezentacje

$$V = \{v_1, v_2, v_3, v_4, v_5, v_6, v_7, v_8, v_9, v_{10}\}$$

$$E = \{\{v_1, v_2\}, \{v_1, v_8\}, \{v_2, v_3\}, \{v_3, v_5\}, \{v_3, v_9\}, \\ \{v_3, v_{10}\}, \{v_4, v_5\}, \{v_4, v_6\}, \{v_4, v_9\}, \{v_5, v_8\}, \\ \{v_6, v_8\}, \{v_6, v_9\}, \{v_7, v_8\}, \{v_7, v_9\}, \{v_8, v_{10}\}, \\ \{v_9, v_{10}\}\}$$

# Reprezentacje

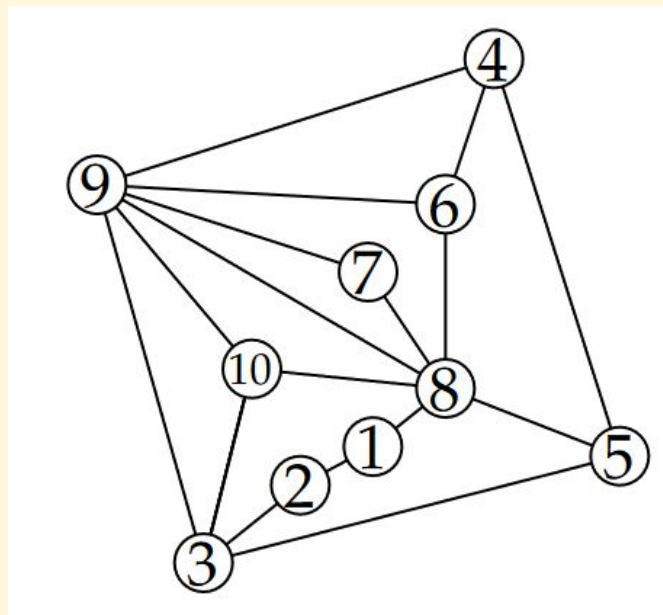
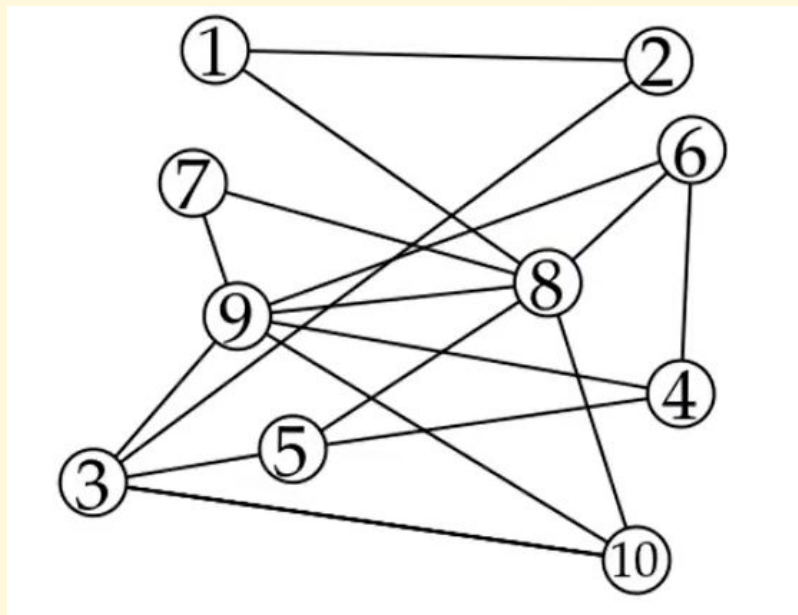
$v_1:$   $v_2, v_8$   
 $v_2:$   $v_1, v_3$   
 $v_3:$   $v_2, v_5, v_9, v_{10}$   
 $v_4:$   $v_5, v_6, v_9$   
 $v_5:$   $v_3, v_4, v_8$

$v_6:$   $v_4, v_8, v_9$   
 $v_7:$   $v_8, v_9$   
 $v_8:$   $v_1, v_5, v_6, v_7, v_9, v_{10}$   
 $v_9:$   $v_3, v_4, v_6, v_7, v_8, v_{10}$   
 $v_{10}:$   $v_3, v_8, v_9$

# Reprezentacje

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

# Reprezentacje

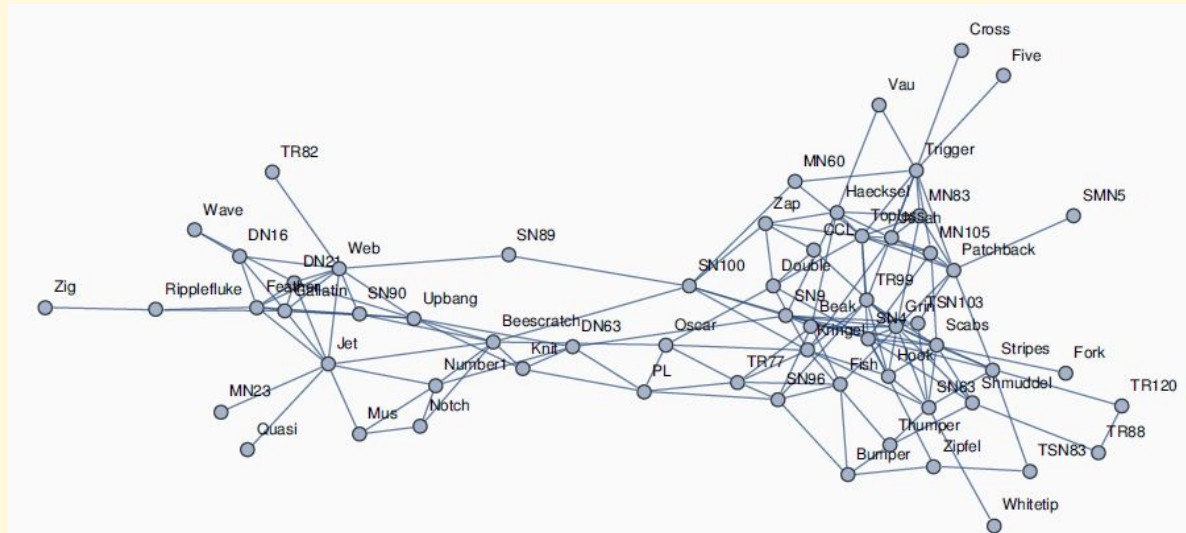


**Grafi są matematyczną reprezentacją  
sieci rzeczywistych i abstrakcyjnych.**



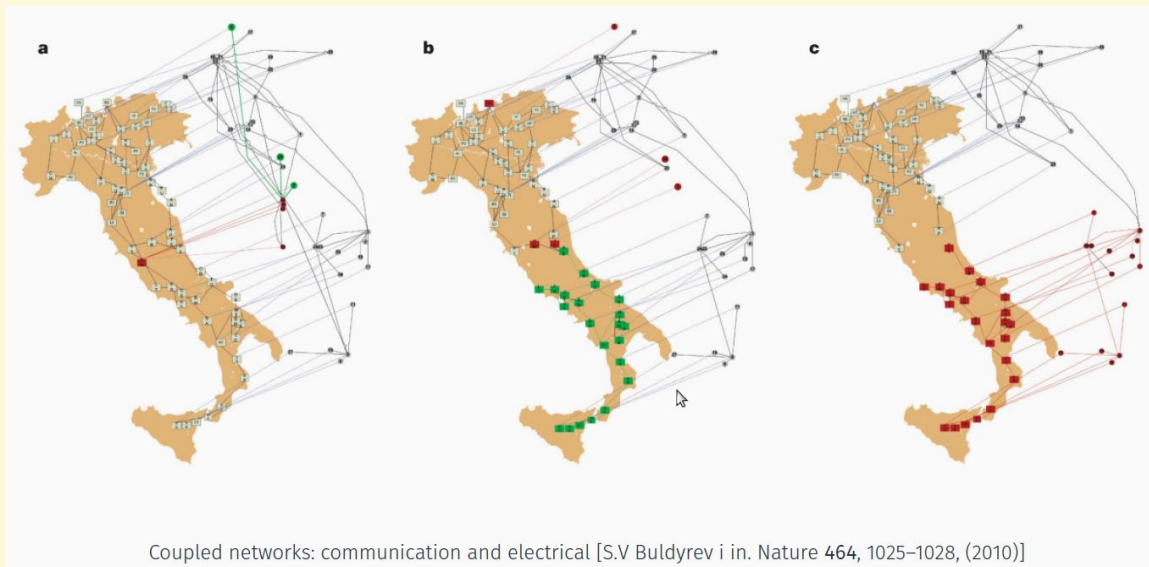
# Grafy abstrakcyjne

- sieć społecznościowa
- sieć komunikacyjna
- sieć filogenetyczna



# Grafy rzeczywiste

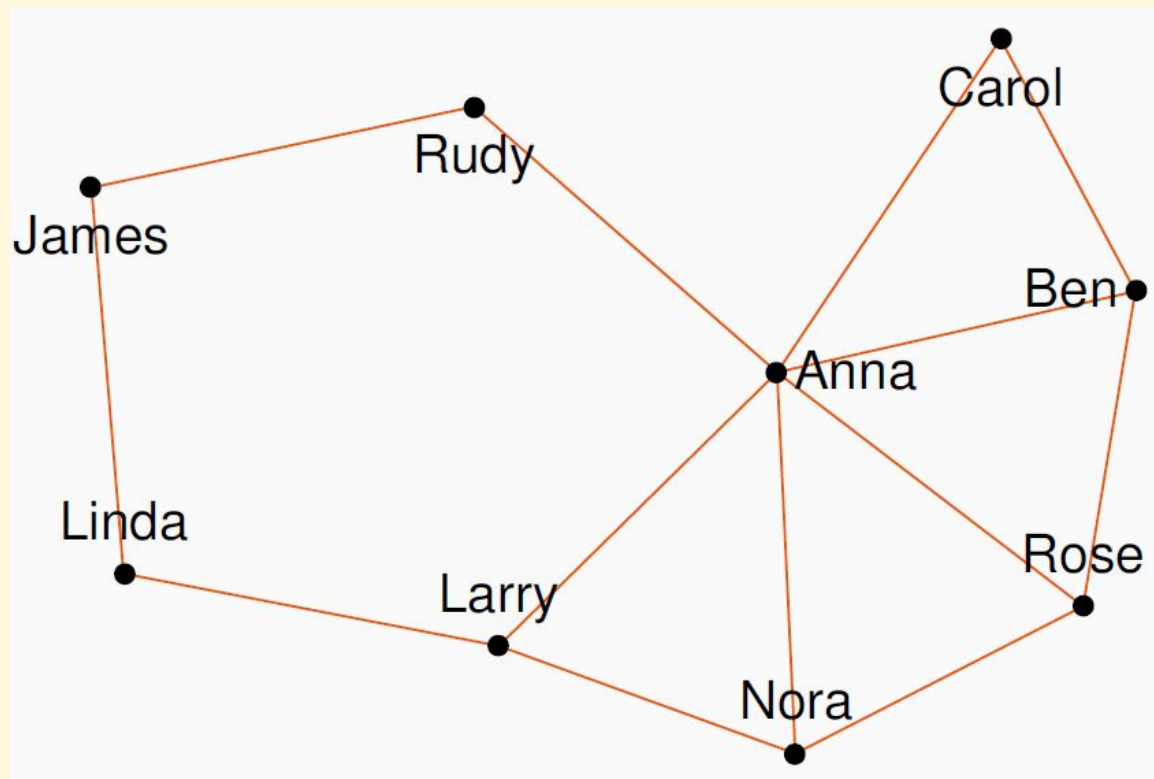
- sieć metra (transportu)
- sieć dróg
- sieć telekomunikacyjna



**Istnieje wiele rodzajów grafów, których podział  
zdeterminowanych jest przez ich strukturę.**

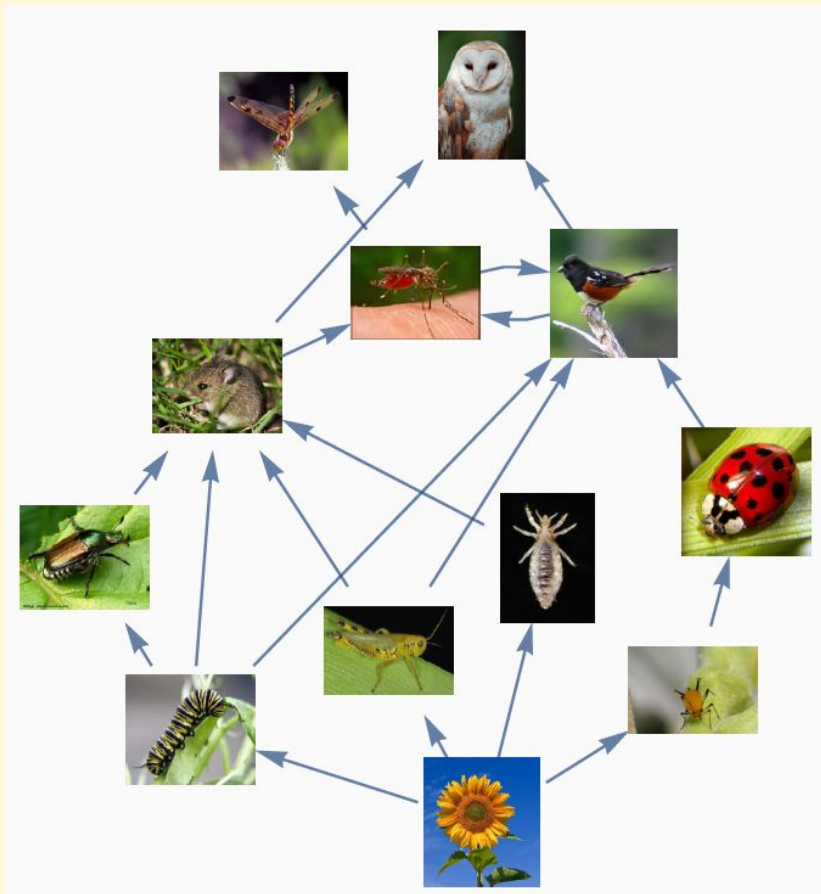
# Rodzaje grafów

- grafy proste



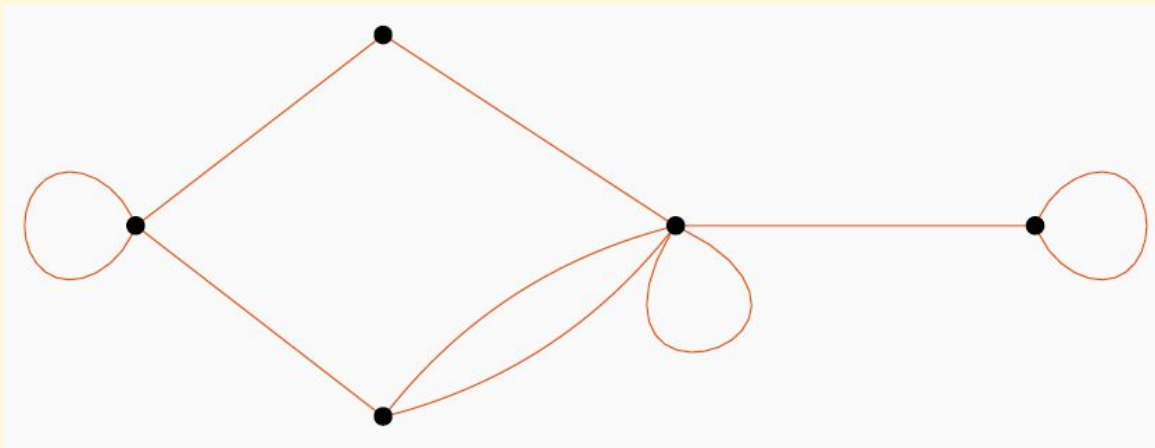
# Rodzaje grafów

- grafy proste
- grafy skierowane



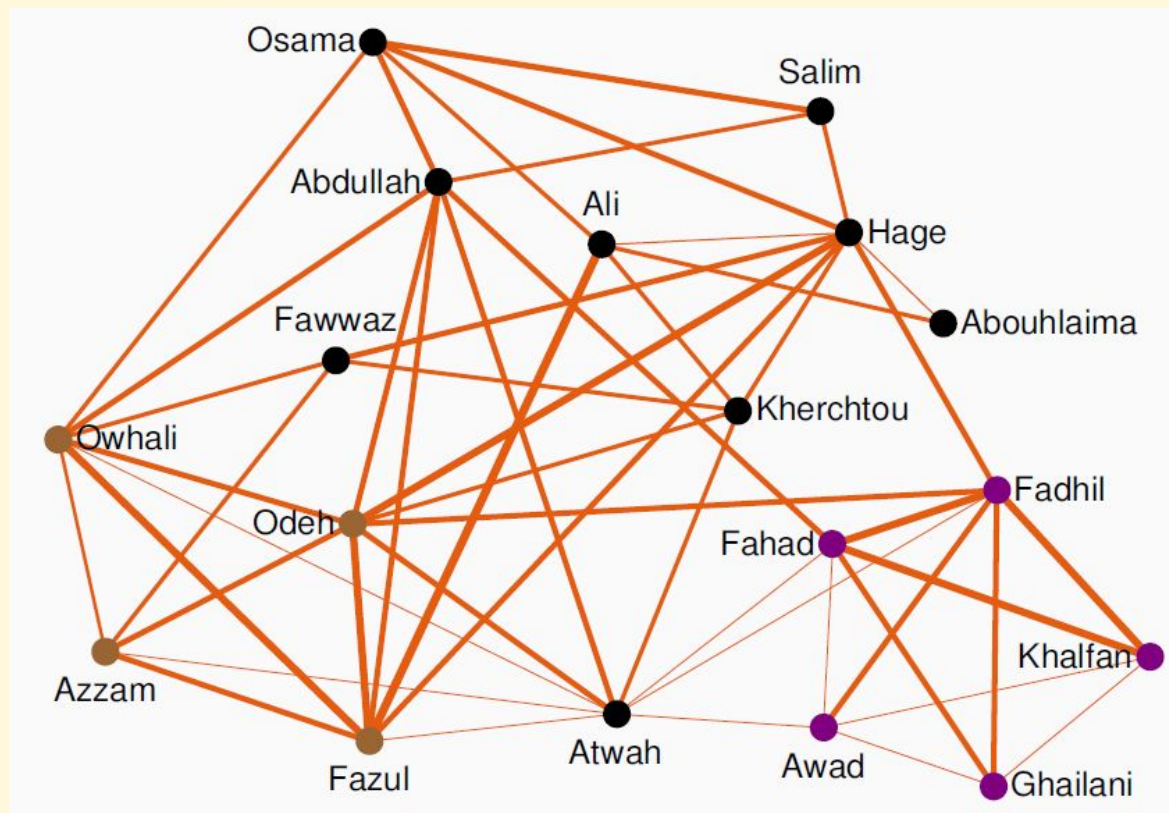
# Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy



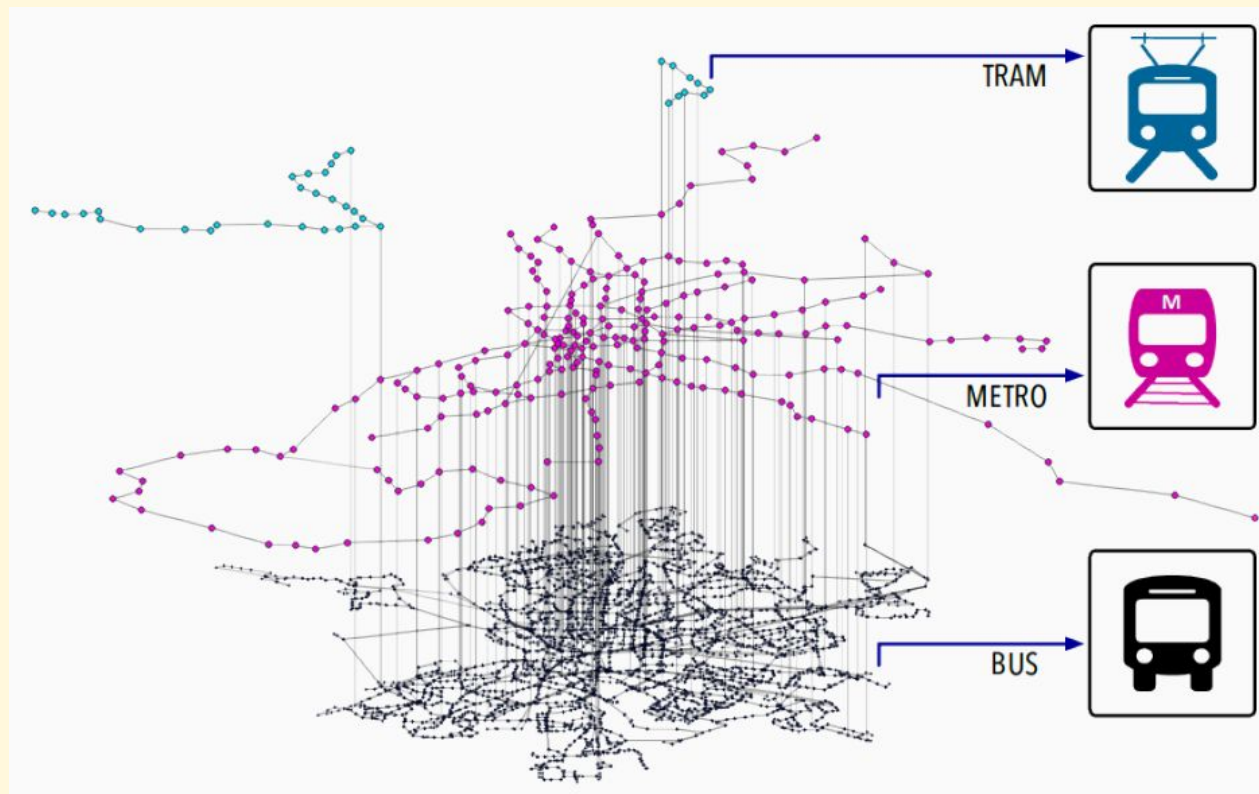
# Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone



# Rodzaje grafów

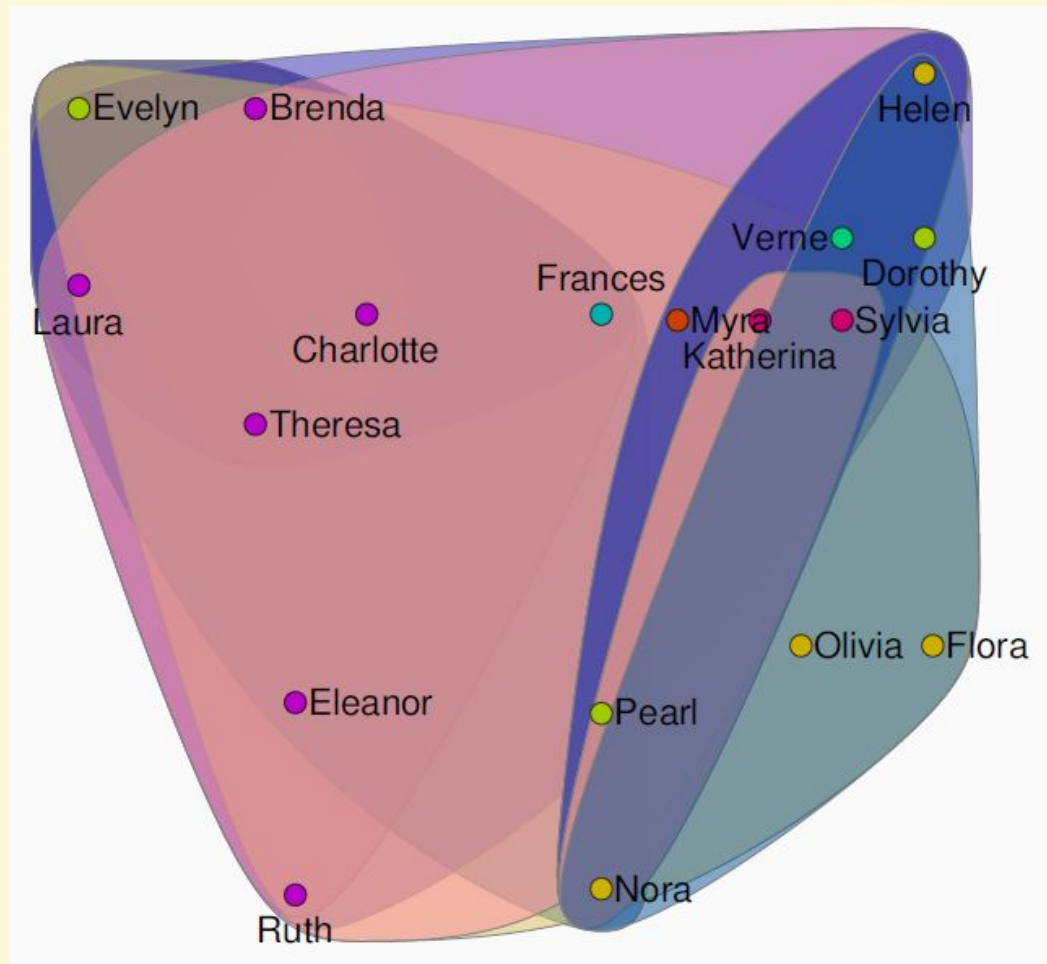
- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone
- grafy warstwowe





# Rodzaje grafów

- grafy proste
- grafy skierowane
- multigrafy
- grafy ważone
- grafy warstwowe
- hipergrafy



1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.
3. Niektóre grafy są zbyt duże, aby rysować je ręcznie.

1. Złożone grafy są trudne do zrozumienia bez dobrej wizualizacji.
2. Wizualizacje pomagają w komunikacji i eksploracji sieci.
3. Niektóre grafy są zbyt duże, aby rysować je ręcznie.

**Potrzebujemy algorytmów, które automatycznie rysują grafy, aby były one bardziej dostępne dla ludzi.**

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.



1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.
4. Często spotykamy grafy dużych rozmiarów, które są trudne do wizualizacji, ze względu na: mnogość wierzchołków, nachodzenie labeli, nakładające się krawędzie, długi czas rysowania grafu.

1. Algorytmiczne rozwiązania potrzebują konkretnych formatów danych.
2. Istnieje wiele formatów w których można przechowywać informacje o grafie (.csv, .json, .gml, .pajek, .dimacs, ...).
3. W większości przypadków nie mamy pełnej kontroli nad dokładnym rozłożeniem wierzchołków.
4. Często spotykamy grafy dużych rozmiarów, które są trudne do wizualizacji, ze względu na: mnogość wierzchołków, nachodzenie labeli, nakładające się krawędzie, długi czas rysowania grafu.

**Wizualizacja grafów jest skomplikowana i nie ma uniwersalnych reguł jak to robić. Niektórych informacji nie da się ukazać samym obrazkiem.**

# Jak mierzyć grafy?

1. Współczynnik klasteryzacji (clustering coefficient)
2. Długość najdłuższej ścieżki / średnica grafu
3. Gęstość sieci ( $2E/N$ )
4. Największy stopień wierzchołka (celebryta)
5. Indeks Hirsch'a
6. PageRank
7. Próg epidemii
8. Liczba Erdos'a
9. Liczba Bacon'a

# Jak mierzyć grafy?

1. Współczynnik klasteryzacji (clustering coefficient)
2. Długość najdłuższej ścieżki / średnica grafu
3. Gęstość sieci ( $2E/N$ )
4. Największy stopień wierzchołka (celebryta)
5. Indeks Hirsch'a
6. PageRank
7. Próg epidemii
8. Liczba Erdos'a
9. Liczba Bacon'a

**Szczegóły na przedmiocie Social Networks & Recommendation Systems - Data Science**

# Liczba Erdos'a

Minimalna ścieżka współautorstwa artykułów prowadząca od naukowca A do Paula Erdos'a. Wartość dla Paula Erdos'a jest równa 0.

MR Collaboration Distance = 3

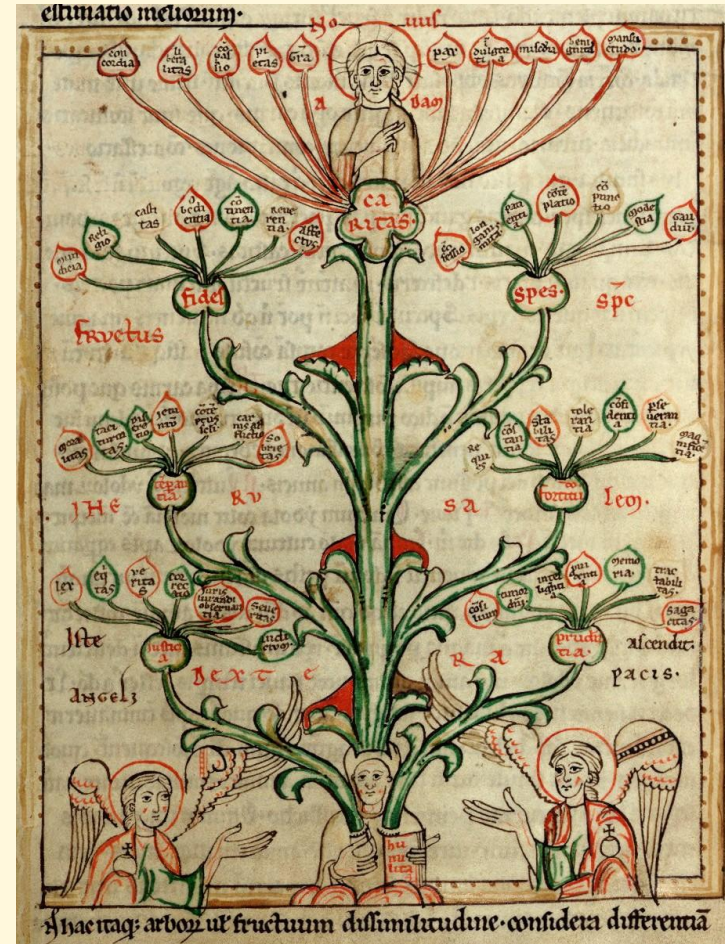
Grzegorzewski, Przemysław	coauthored with	Mesiar, Radko	MR3660830
Mesiar, Radko	coauthored with	Širáň, Jozef	MR4171596
Širáň, Jozef	coauthored with	Erdős, Paul <sup>1</sup>	MR1297187

MR Collaboration Distance = 4

Cena, Anna	coauthored with	Gągolewski, Marek	MR4240238
Gągolewski, Marek	coauthored with	Mesiar, Radko	MR3158692
Mesiar, Radko	coauthored with	Širáň, Jozef	MR4171596
Širáň, Jozef	coauthored with	Erdős, Paul <sup>1</sup>	MR1297187

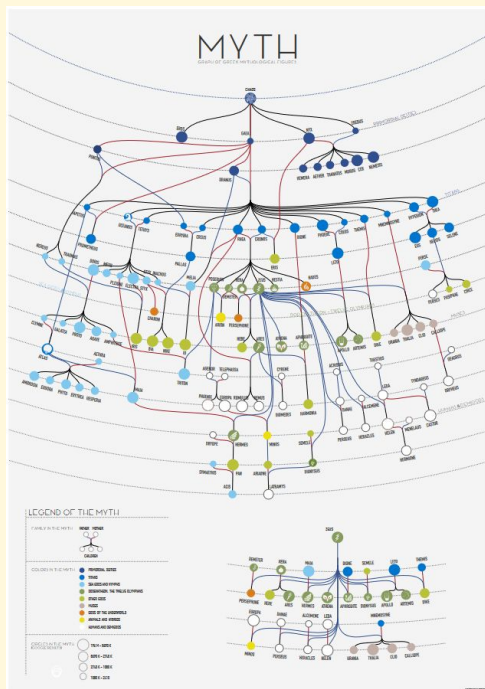
# Drzewo cnót głównych

Cztery cnoty kardynalne umieszczone są niżej niż trzy cnoty teologiczne. Z każdej z cnót głównych wyrastają cnoty niższego rzędu. Speculum Virginum, XII w.

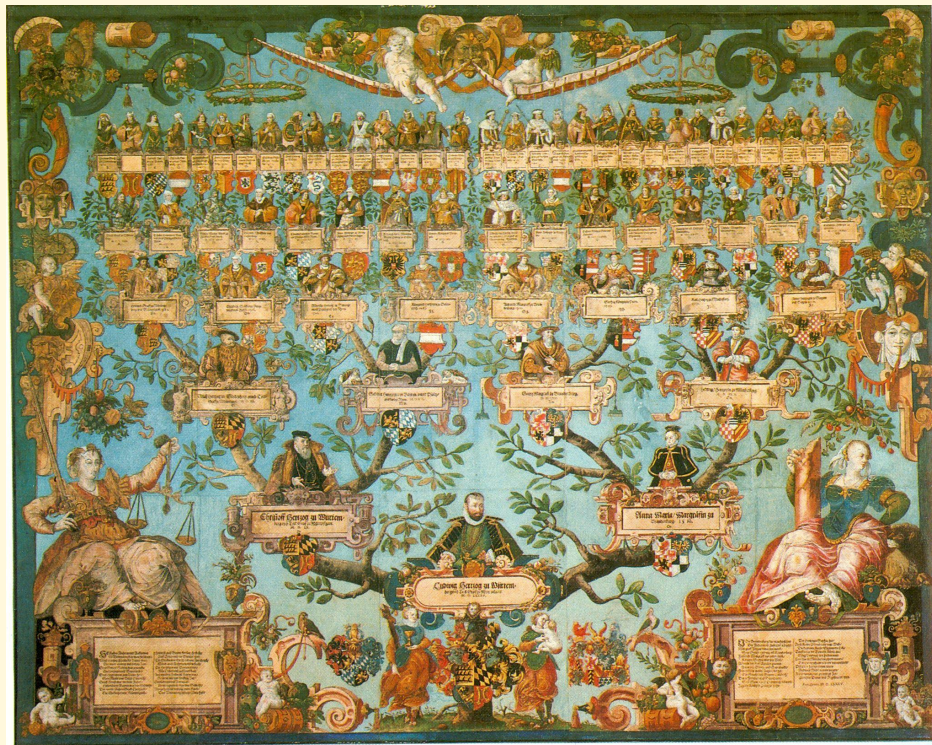




# Drzewo genealogiczne (sieć społecznościowa)

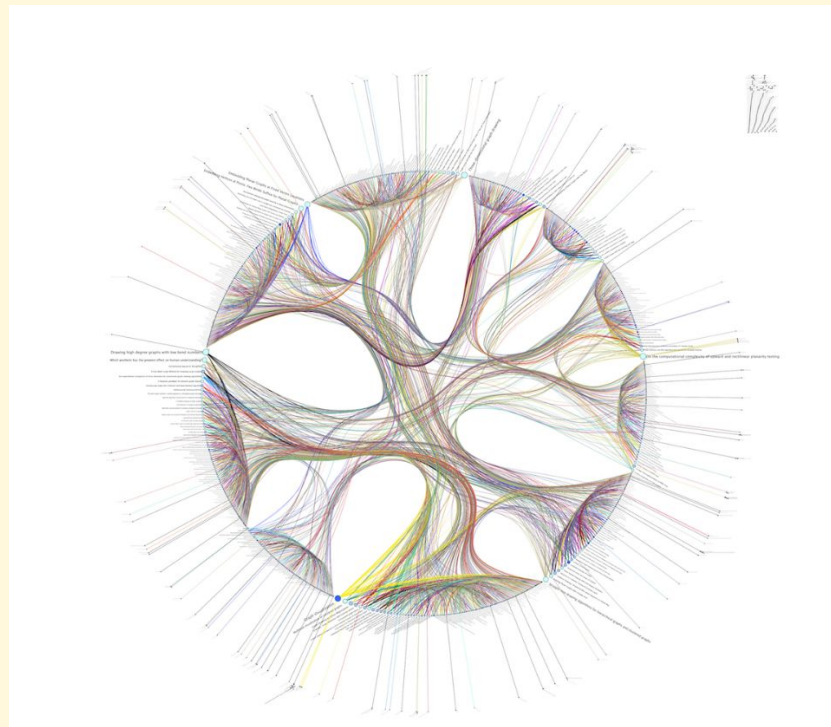


J. Klawitter, T. Mchedlidze, Link: [go.uni-wue.de/myth-poster](http://go.uni-wue.de/myth-poster)



Ahnentafel Herzog Ludwig von Württemberg, 1585

# Graf cytowań (sieć społecznościowa)



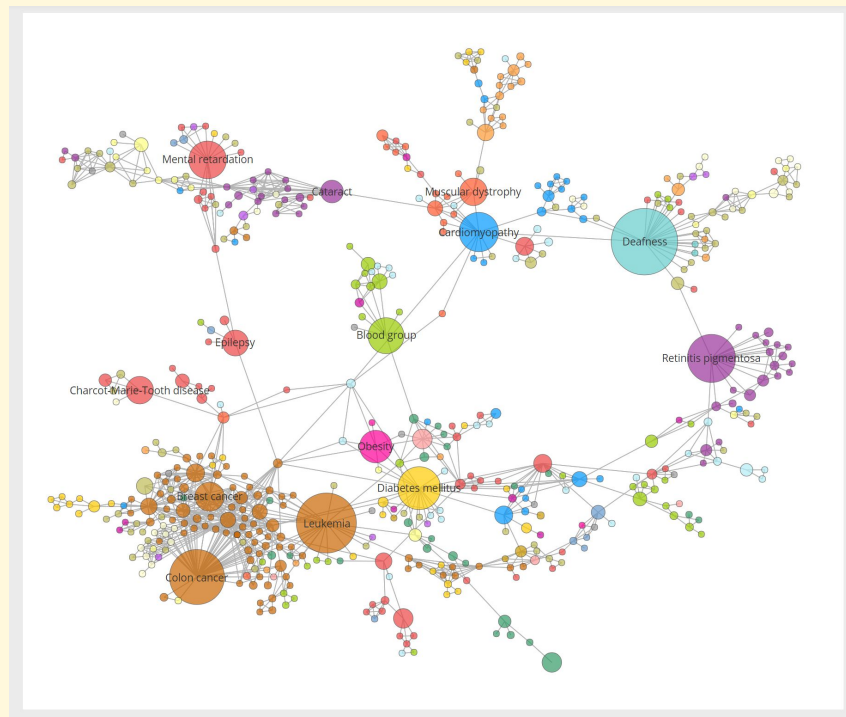
Da Ye, Link: <https://go.uniwue.de/citation-graph>



# Kolej dużych prędkości w Europie (sieć komunikacyjna)

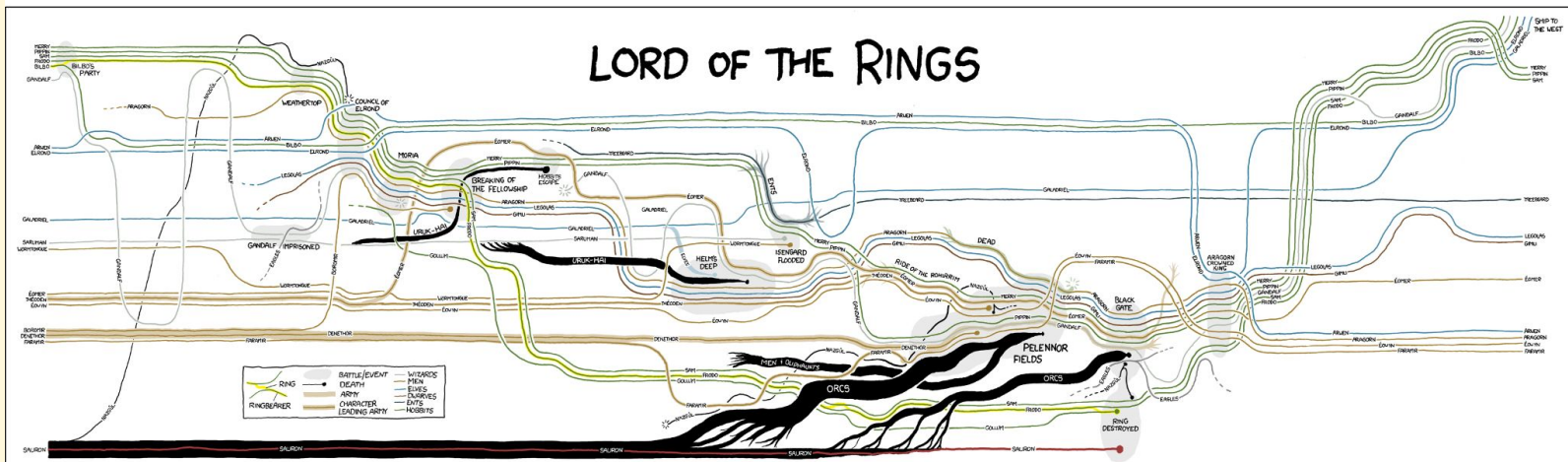


# Sieć ludzkich chorób (bioinformatyka)

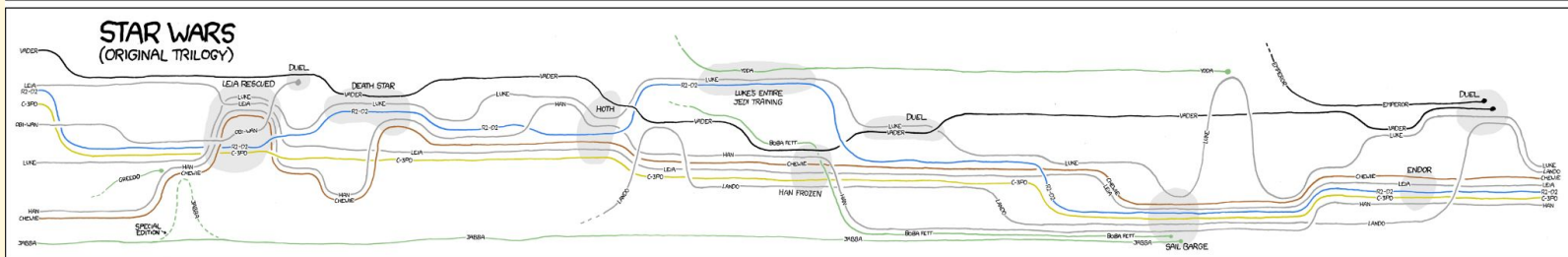


THESE CHARTS SHOW MOVIE CHARACTER INTERACTIONS.  
THE HORIZONTAL AXIS IS TIME. THE VERTICAL GROUPING OF THE  
LINES INDICATES WHICH CHARACTERS ARE TOGETHER AT A GIVEN TIME.

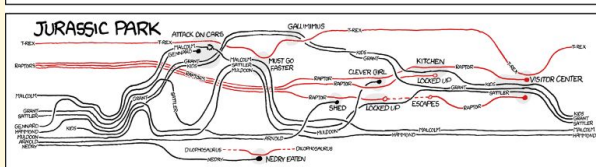
## LORD OF THE RINGS



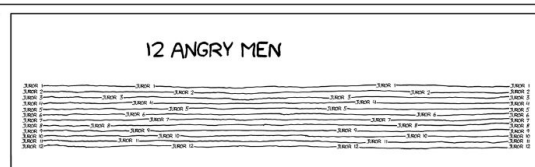
## STAR WARS (ORIGINAL TRILOGY)



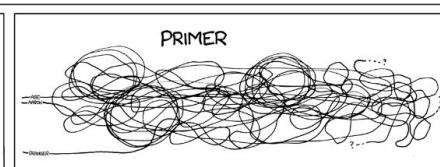
## JURASSIC PARK



## 12 ANGRY MEN



## PRIMER



# Więcej o grafach

- <https://seafire.rlp.net/f/f36d7e005a3c48a2bac2/>
- Przedmiot Social Networks & Recommendation Systems na studiach magisterskich Data Science - dr. inż. Grzegorz Siudem
- The Anatomy of a Search Engine: Sergey Brin and Lawrence Page