Wrocław University of Technology

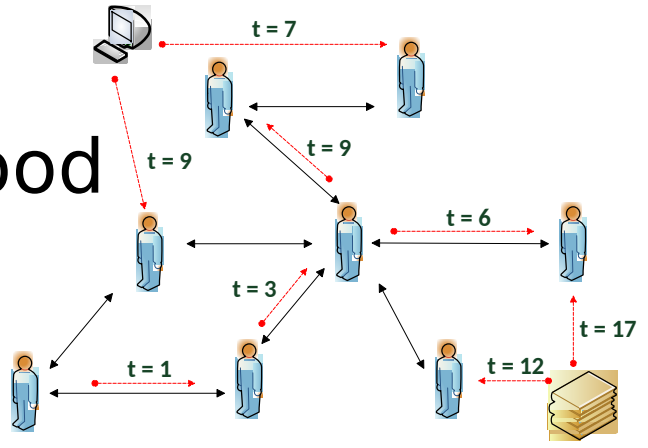
Computational Social Science

Radosław Michalski

Data Science course

My Areas of Interest

- Diffusion of information
  - a general idea o spreading the information within the network
- Diffusion of innovations
  - ideas, products, behavior
  - observing the neighborhood
- Social influence
  - direct influence of others on an individual



My Areas of Interest - Applications

- Social networking websites
  - finding and predicting the trends
  - finding important members (influencers)
- Marketing/social campaigns
  - how to promote products/ideas?
- Enterprises
  - maximizing the spread of information (broader/faster)
- Blockchain
  - analysing network of transactions

How to Find me?

https://www.ii.pwr.edu.pl/~michalski

radoslaw.michalski@pwr.edu.pl

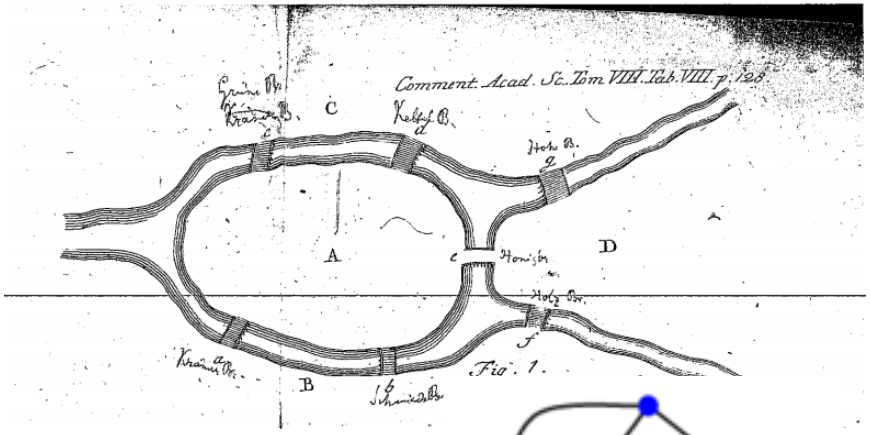
https://github.com/rmhere

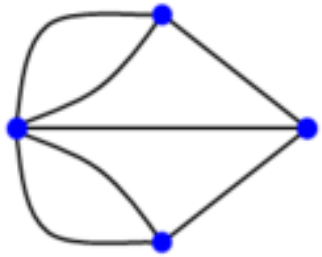
Wrocław University of Technology

Graph and Networks

The Story - Koenigsberg Bridges

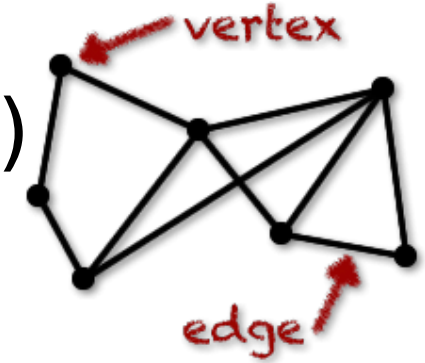
- Koenigsberg is now a part of Russia (Калининград)
- A river Преголя has two islands there
- Is it possible to cross all seven briges once?
- Leonhard Euler proved that it is impossible in 1741 – the origins of graph theory





Graph Theory – Basics

- Graph as a basic entity
- It consists of nodes (vertices) and links (edges)
- Different (basic) scenarios:
  - vertices can be labelled or not
  - edges can be weighted or not
  - edges can be directed or not



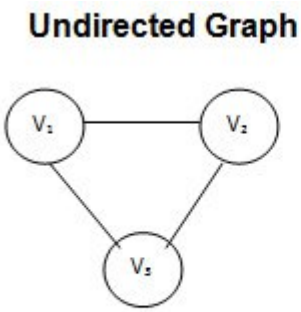


Figure 1: An Undirected Graph

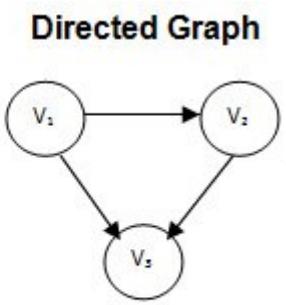
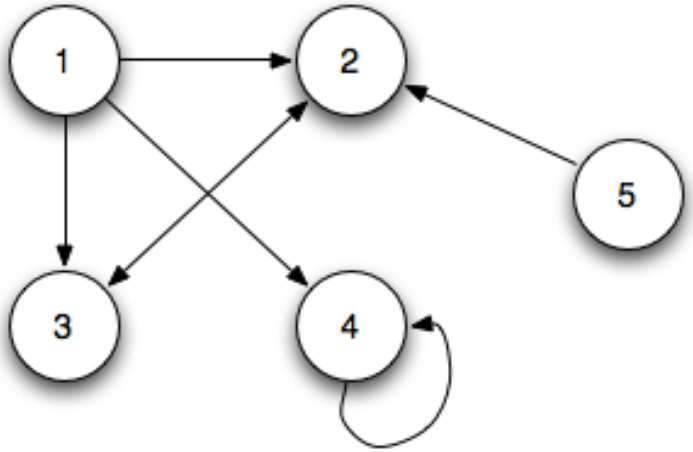


Figure 2: A Directed Graph

Representing Graphs

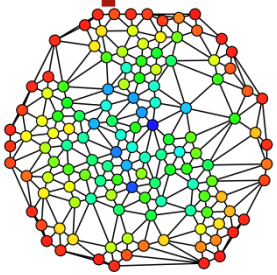
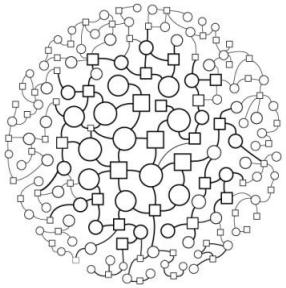
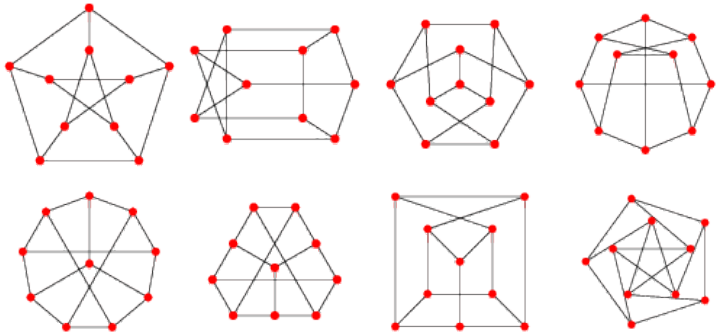
Visual representation vs. **adjacency** matrix



	1	2	3	4	5
1	0	1	1	1	0
2	0	0	1	0	0
3	0	1	0	0	0
4	0	0	0	1	0
5	0	1	0	0	0

From Graphs to Networks

- Graph is a **mathematical representation**

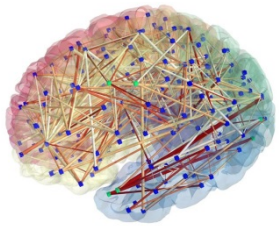




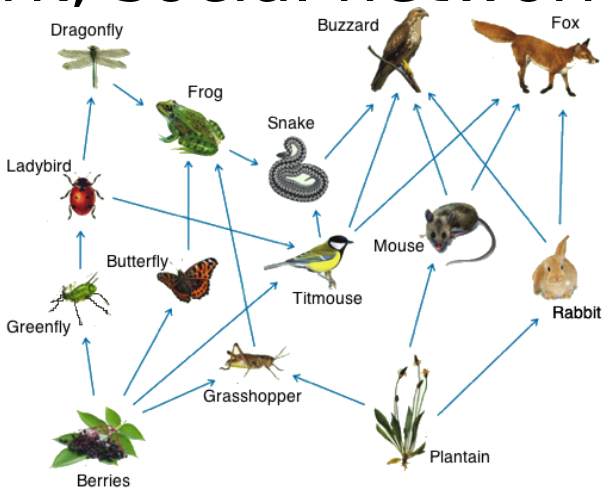
- Network is linked to **real system**
  - web
  - human brain
  - social structure
  - power grids




Kinds of Networks

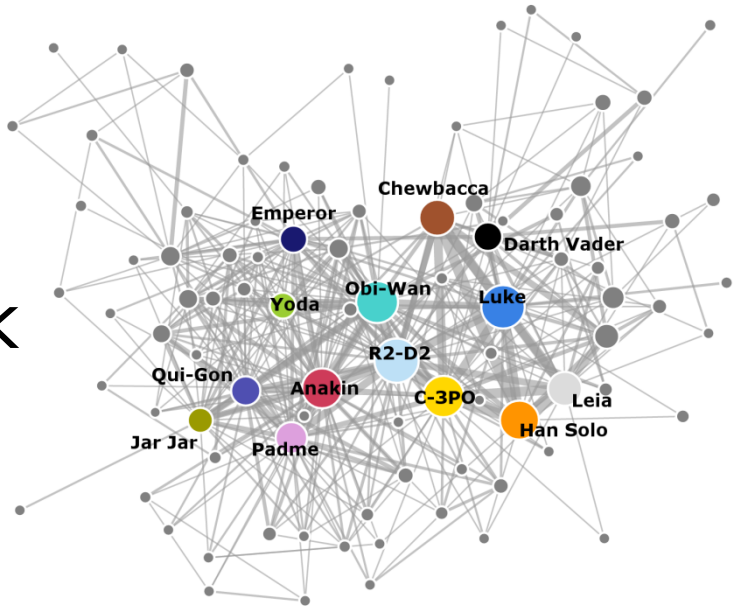
- **Social networks**
  - represent relationships between humans
  - e.g. corporate social network, social network of twitter users, ...
- **Biological networks**
  - human brain
  - food web
- **Communication networks**
  - structure of the web





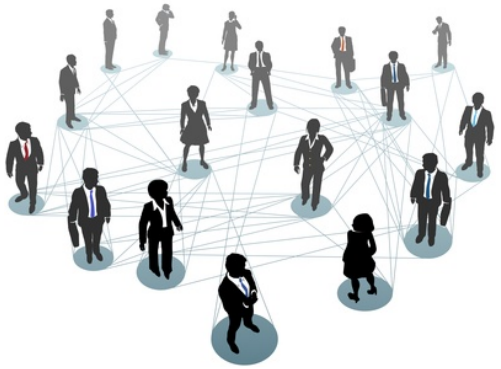
How Networks are Different from Graphs?

- Typically the **vertices are labelled**, since they link to real objects
- Network represents a **part of reality**, not the whole world:
  - social network of students
  - people using Facebook
  - power grid in Europe



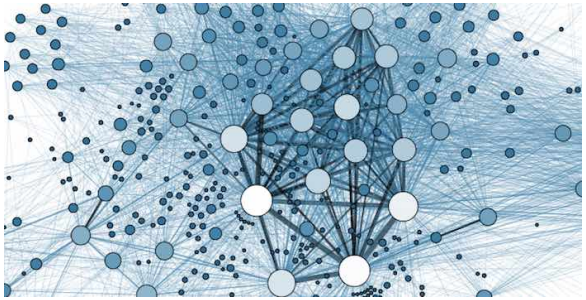
Networks – Vertices

- Vertices typically have some more **attributes** than just labels
  - employees social network:
    - gender
    - position
    - work experience
  - power grids
    - load/capacity
    - location

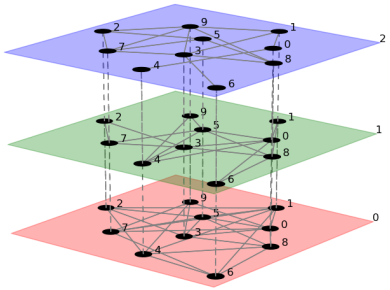


Networks – Edges

- Edges are typically **weighted**:
  - volume of communication
  - length of relationship
  - sales' value



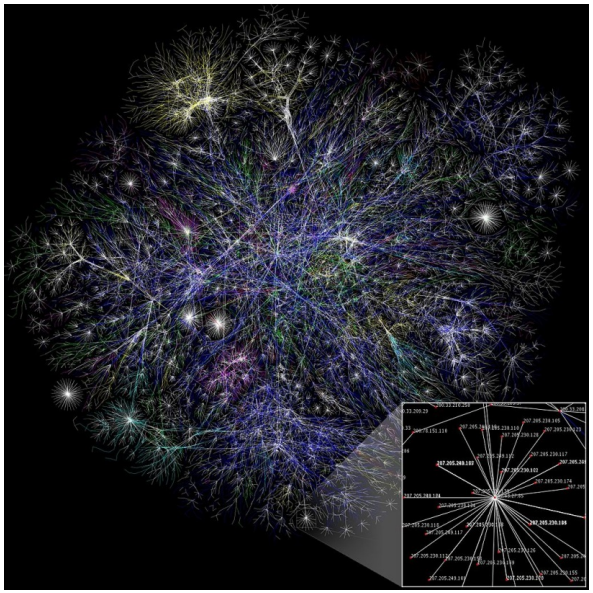
- Different edges' kinds – **multilayer** networks
  - type of communication (mail, phone, face to face)



Network Science I

- **Graphs** – mathematical representation
- **Networks** – real world objects

- **Network science** => using the knowledge for solving specific real-world problem



Network Science III

- Real world problems – what are they?
  - what **employees** are crucial for the company?
  - can we loose this **mobile phone user**?
  - if this **brain stucture** will be damaged, will this person still conscious?
  - which **banks** can’t bancrupt?
  - if he/she will get **infected**, who else will?
  - if this part of **power grid** will fail, which parts will fail as well?
  - who should receive an **iPhone** to convince the most of the classmates to buy the same phone

Network Science - Future

- **Big Data** – huge datasets
- **Computational challenges** – computing clusters etc.
- **Confirming** the models in data or finding new models in data (e.g. human behaviour)
- Summing up: data-driven approach

Wrocław University of Technology

What is a social network?

Social Network

- Typically, social network is a network where vertices are people and links represent relationships.
  - vertices: actors, people
  - links: relationships, interactions

- Formally, it is still a graph
  - SN = <V, E>

Social Network – Examples

- **Examples** of social networks:
  - corporate social network
  - social network of freshmens at the university
  - social network in neighbourhood
  - social network of twitter users
- **Social network** is not a **social networking website (socia media)**
  - yet, by using them, we form social networks

Social Network Analysis

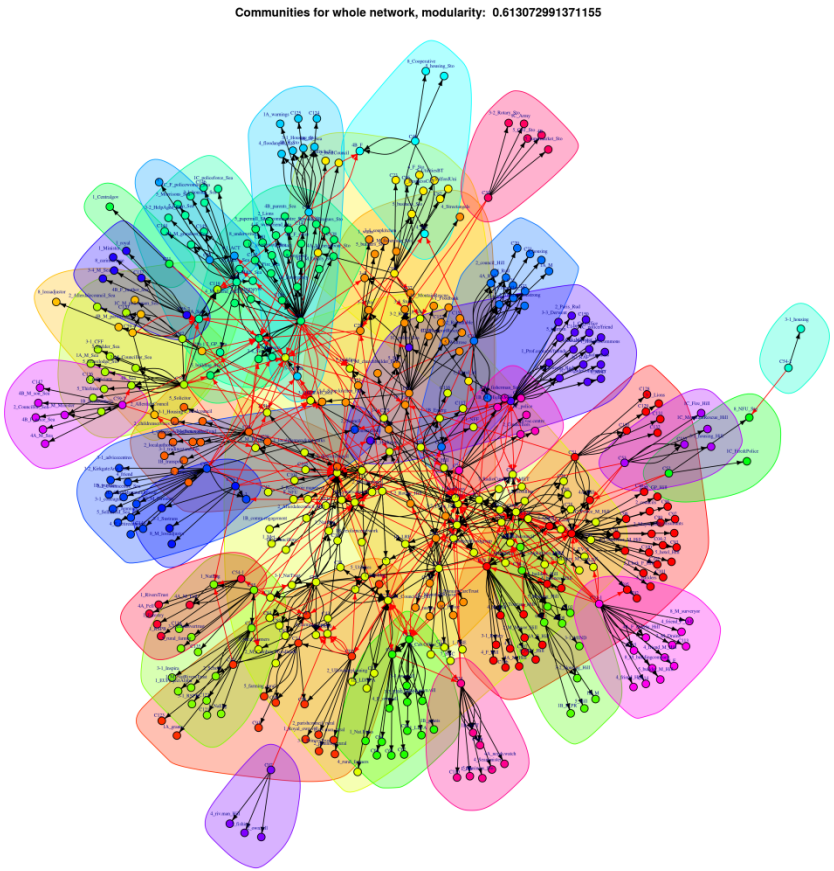
This is a right time to analyse social network: **Social Network Analysis (SNA)**

SNA is a set of tools and techniques to analyse social networks giving us **deeper understanding** of the network structure.

What are the **crucial methods** of SNA?

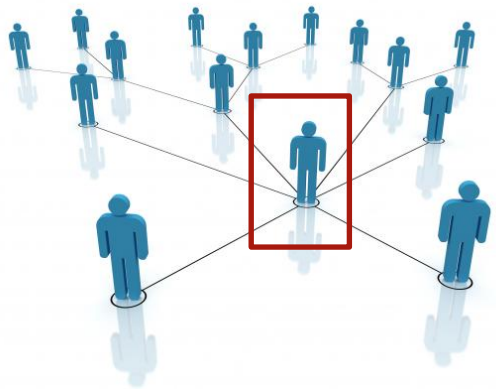
Areas of SNA

- **Local level**
  - analysing nodes and links
  - position, centralit
- **Network level**
  - global network characteristics
- **Group level**
  - groups and clique



Analysing Nodes I

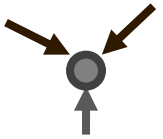
- What can we say about node in the network?
  - is it connected with others?
  - how well is it connected?
  - is it centrally positioned?
  - does it have common neighbours with its neighbours?



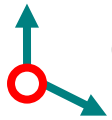
  - any more ideas?

Degree and Centrality

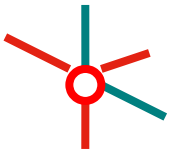
- Node properties
  - from direct connections
    - indegree
      how many directed edges (arcs) are incident on a node
    - outdegree
      how many directed edges (arcs) originate at a node
    - total degree (in or out)/degree
      number of edges incident on a node
  - from the entire graph
    - centrality (betweenness, closeness)



indegree=3



outdegree=2



degree=5

Degree - Properties

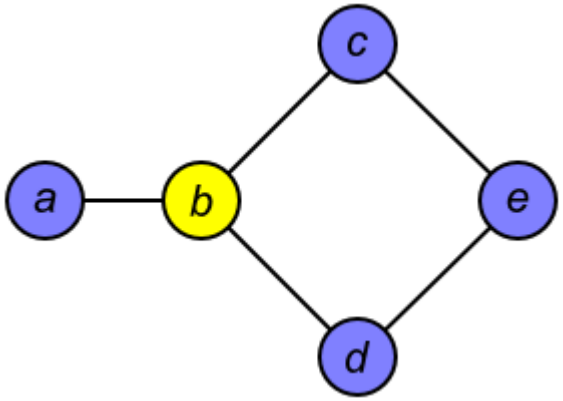
- A node can be well connected to its neighbours but in the periphery of the network
  - information originating there is distributed slowly across whole network
  - initially it reaches many neighbours, but then it has long way to go

How to Find More Central Nodes?

- Centrality is related to the **whole structure**, not the node itself.
  - maybe not the number of adjacent links matters?
  - if located more centrally, I can spread information faster than the red node?
  - how to quantify that?

Betweenness Centrality

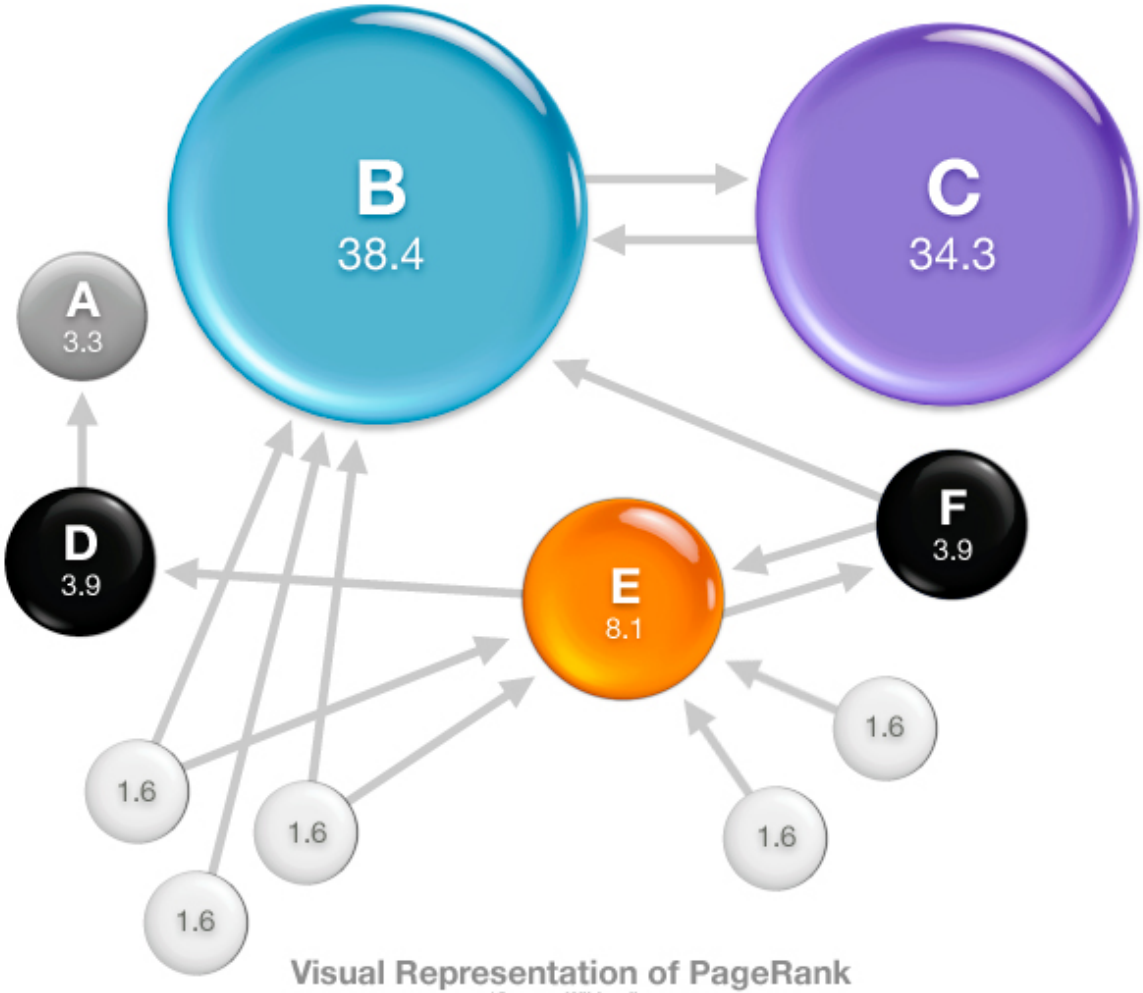
- Compute shortest paths going from any two nodes but not ***b***



- ***On how many of those b is present?***

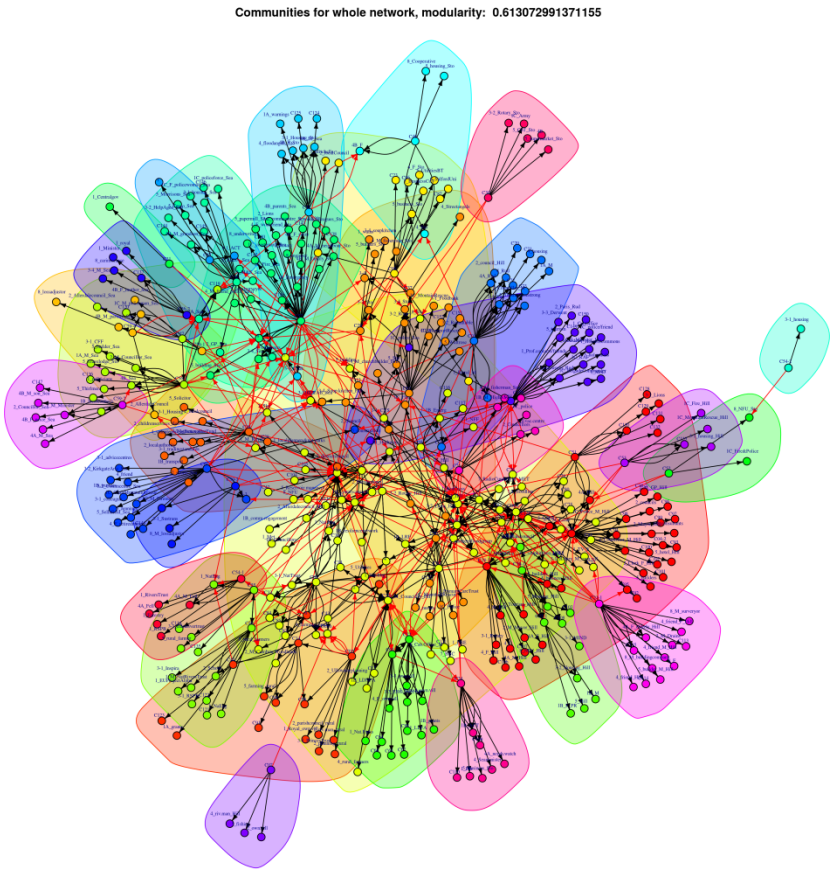
$$g(v) = \sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$$

PageRank


B
38.4
C
34.3
A
3.3
D
3.9
E
8.1
F
3.9
1.6
1.6
1.6
1.6
1.6
Visual Representation of PageRank
*Source: Wikipedia.

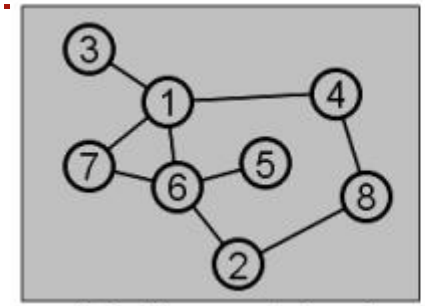
Areas of SNA

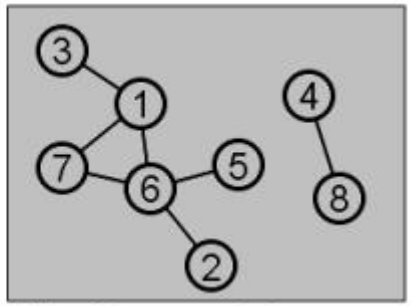
- **Local level**
  - analysing nodes and links
  - position, centralit
- **Network level**
  - global network characteristics
- **Group level**
  - groups and clique

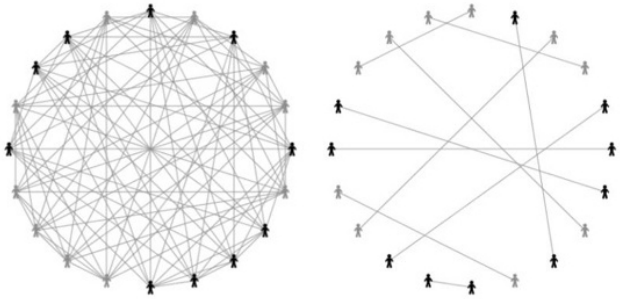


Network Level

- Global characteristics:
  - **diameter** – the longest shortest path
  - **average path length** – the averaged shortest paths
  - **sparsity** – ratio of links and nodes (the more zeros in the adjacency matrix, the sparser the network)

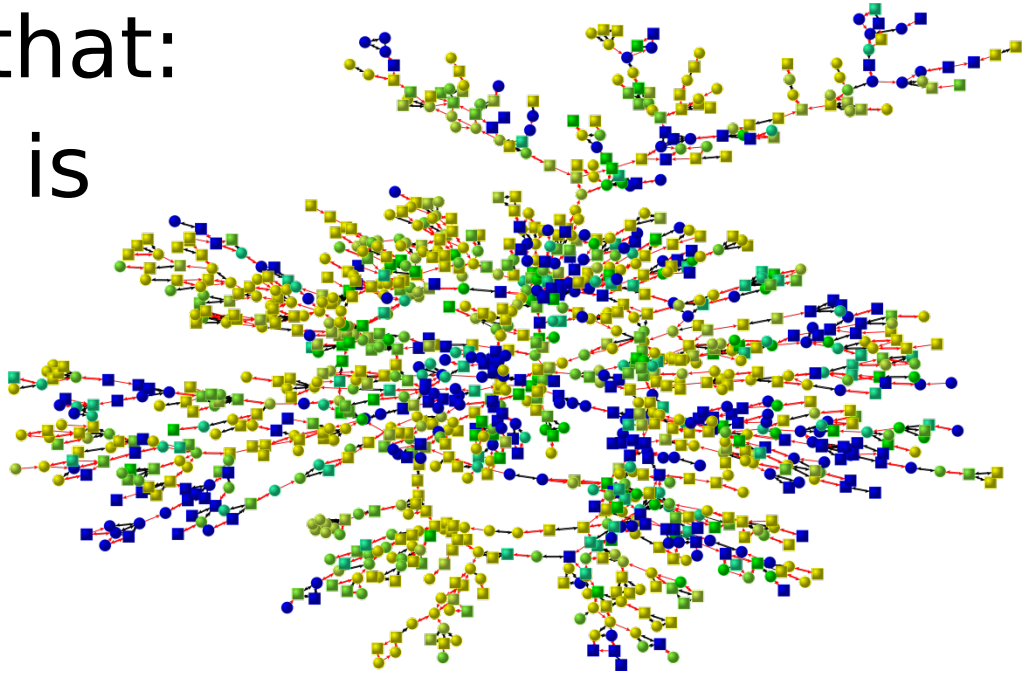


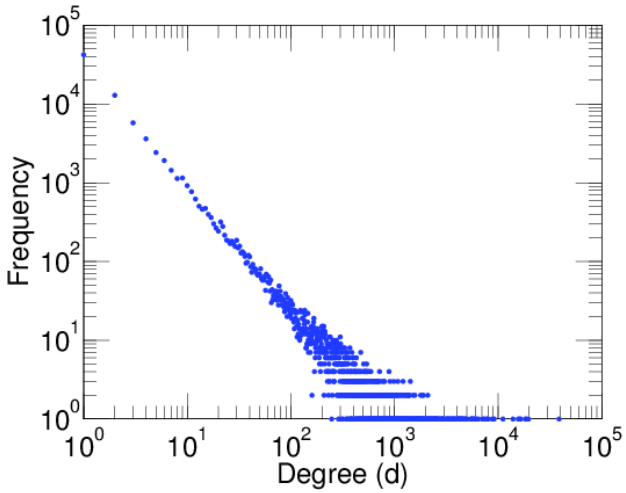




Distributions in social networks

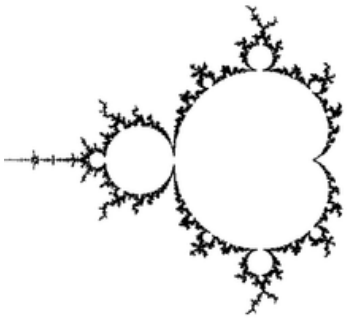
- As the network is typically like that:
- The distribution is like that:



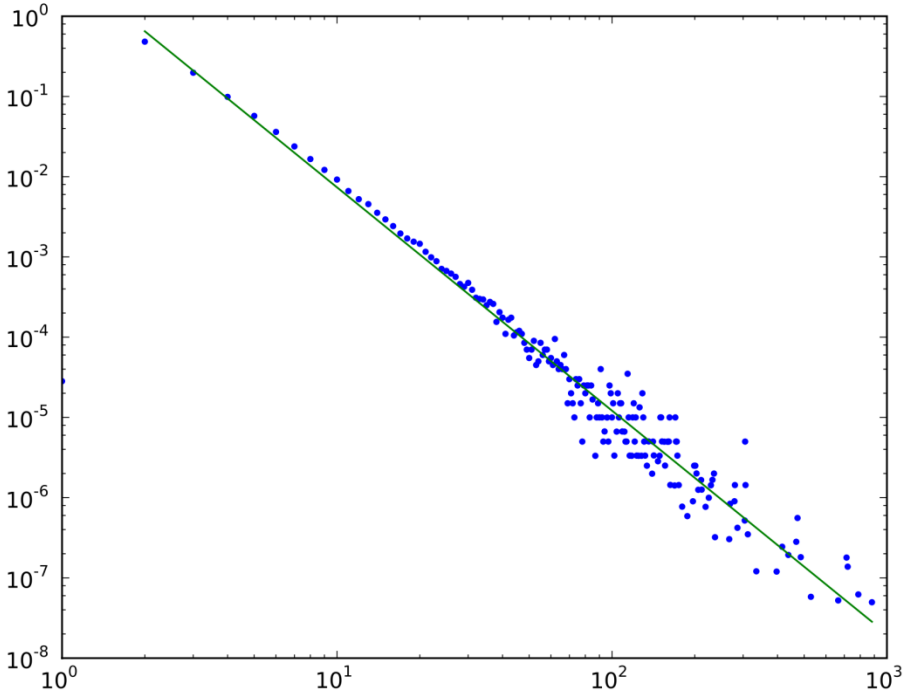


Power law in networks

- **Scale-free** phenomenon (log-log scale)
- Few nodes that are **order of magnitude** better connected than others
- No particular scale -> fractals

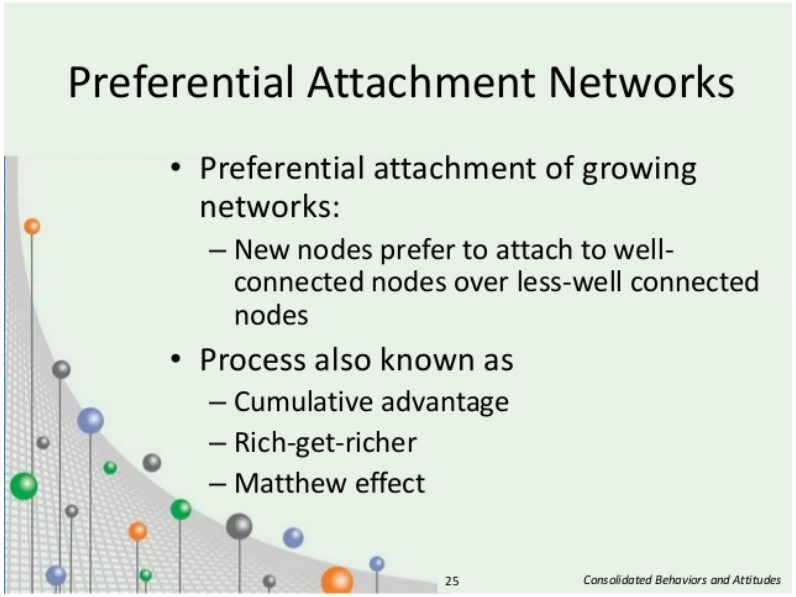


- **Long tail**



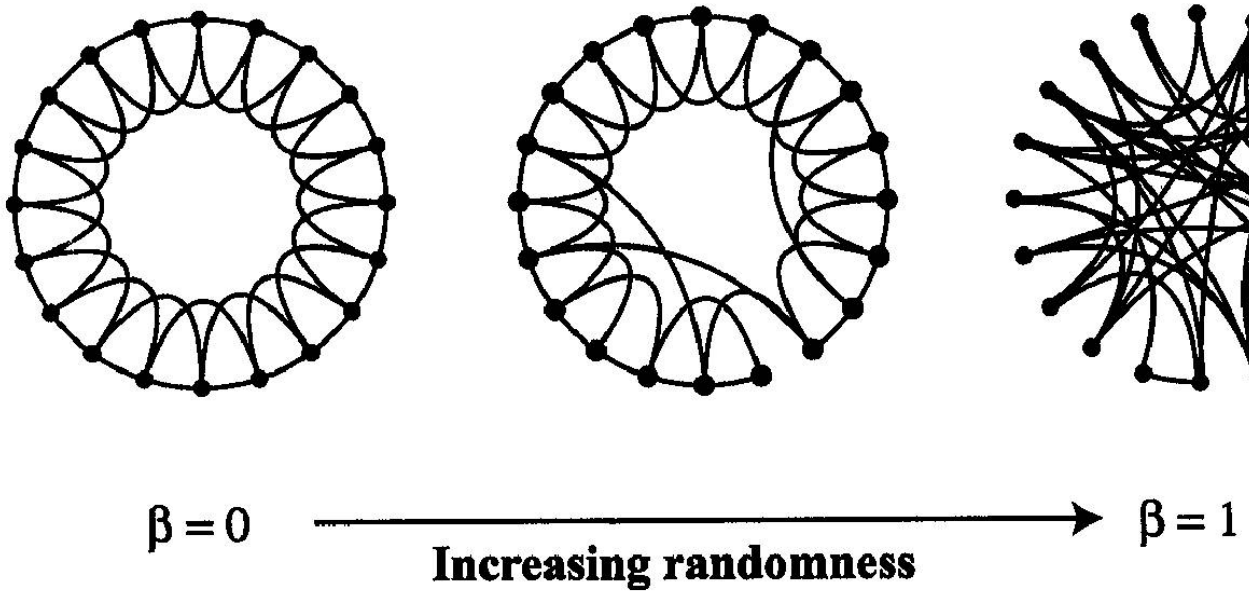
Preferential attachment model

- When joining a social networking website (e.g. twitter) **who will be followed by you**?
  - famous and popular ones
  - regular twitter users
- **Rich gets richer** phenomenon
- **Power law** works!



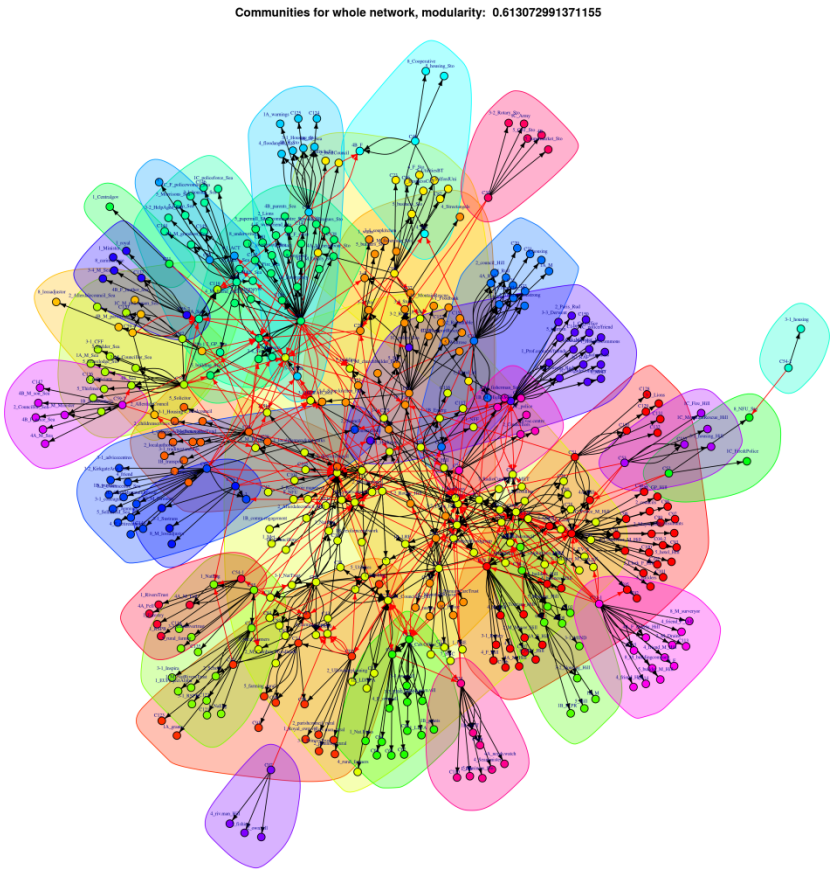
Small world model (Watts Strogatz)

- **Some people** are **better connected** then most of others



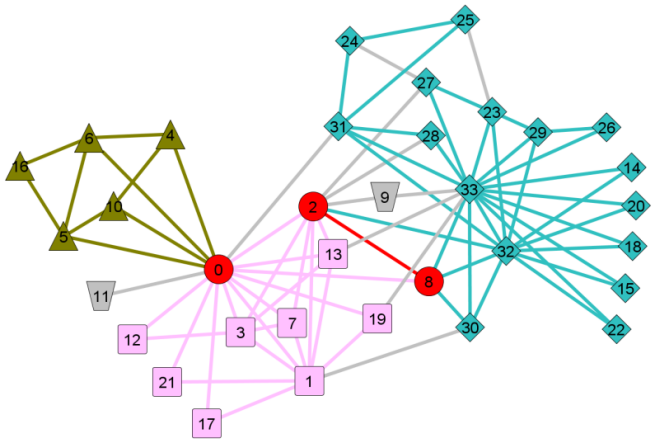
Areas of SNA

- **Local level**
  - analysing nodes and links
  - position, centralit
- **Network level**
  - global network characteristics
- **Group level**
  - groups and clique



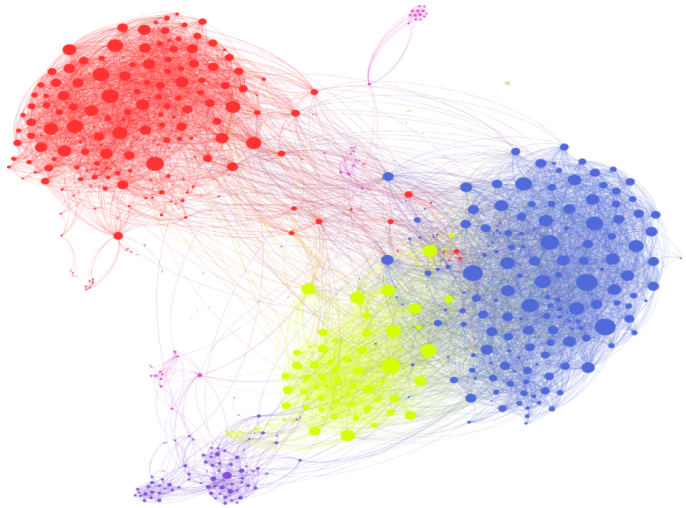
Do we Form any Groups?

- If we analyse **Facebook** data, do you think that people contact there at random or form some **communities**?
  - regular contacts with the same people,
  - less random interactions,
  - stability over time.
- If so, how to detect groups?



Group Detection

- The easiest – **attributes of vertices** (if we have them) – same class, same town etc.
- Typically, we don't have this data, so we use **algorithms**
  - **a group** represents a structure that is **more connected within** than outside



What about labs?

- Software tools for SNA (and much more)
  - R
  - Rstudio
  - igraph

- Applications of network science
  - how to make use of the knowledge we have

Courses / books

On-line courses

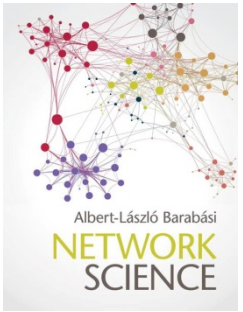
- **Coursera: Social Network Analysis**
  (Lada Adamic, see youtube)
- „**Introduction to Network Science**” Leonid Zhukov (youtube)

Books

- Albert-Laszlo Barabasi „**Network Science**”
- Wasserman and Faust „**Social Network Analysis**”

Keywords

social network analysis, network science, complex networks, network models



Materials for data sciencists

Books

- Przemysław Biecek „**Przewodnik po pakiecie R**”
- Grolemund and Wickham „**R for Data Science**”)available online here: http://r4ds.had.co.nz/)
- Bruce and Bruce „**Practical Statistics for Data Scientists: 50 Essential Concepts**"

Courses

- Coursera data science courses:

  https://www.coursera.org/browse/data-science?languages=en