Complex networks

Basic graph theory, adjacency matrix, distance, path, connectedness, clustering

János Török

Department of Theoretical Physics

February 28, 2023

Information

- Coordinates:
  - Török János
  - Email: torok.janos@ttk.bme.hu, torok72@gmail.com
  - Consultation:
    - F III building, first floor 6 (after the first stairs to the right, at the end of the corridor), Department of Theoretical Physics
    - Upon demand (Email)
- Webpage: https://physics.bme.hu/BMETE15MF76_kov?language=en
- Homework: http://edu.ttk.bme.hu/

Requirements

- Signature
  - 40% from each homework
- Exam: mark
  - 50%: From homeworks (individual)
  - 50%: From projects (individual, pairs) presented in the last lecture, or exam
  - Turn it in language: English, Hungarian, German, French

Subjects

Lecture is based on the lectures of János Kertész

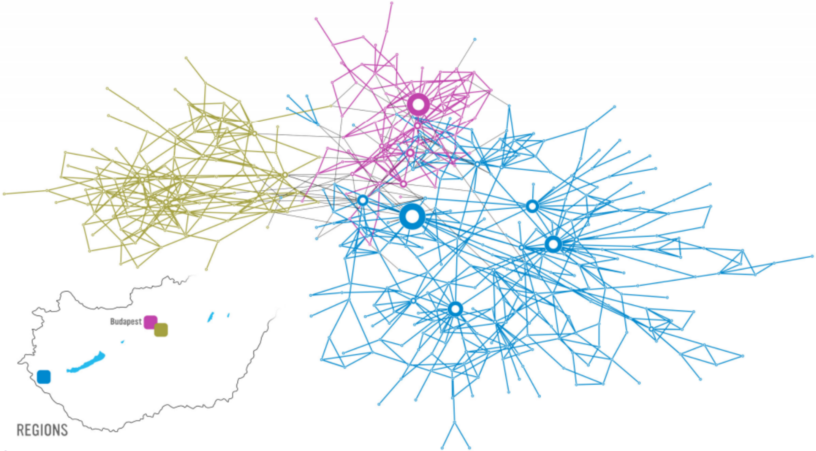
1. Complex Networks: graphs with non-trivial features, graph theory
2. Basic andom network models (Erdős-Rényi, Watts-Strogratz)
3. Preferential attachment, scale free networks, configuration model
4. Stochatic block model
5. Growth models and cascades
6. Temporal networks
7. Diffusion on networks
8. Robustness and spreading
9. Communities
10. Core-periphery
11. Hierarchy
12. Sampling
13. Navigation on networks

Introduction

- Complex Networks: graphs with non-trivial features
    - Networks: graphs, which are nodes and edges
    - Graphs: Objects with interactions
    - Hope: network structure can help us understand the system

Example (my favourite)

- Hungarian company 3 bases



Maven 7 from `networksciencebook.com` by Barabasi.

Example (my favourite)

- CEO (red), top managers (blue), Managers (magenta), group leaders (orange)

Example (my favourite)

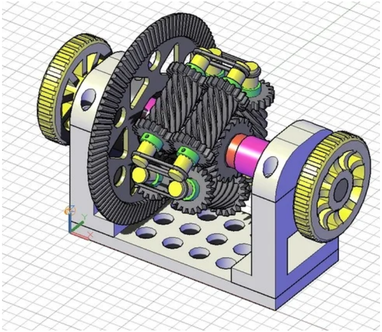
- Biggest hub, and links at distance 1 and 2

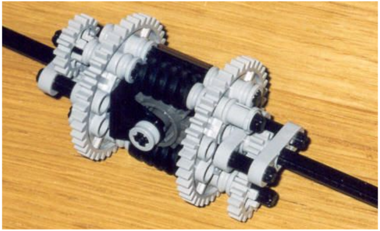
Complex networks

- Social connections
- IT connections
  - Hardware
  - WWW
- Biology
  - Food web
  - Metabolism
  - Neural connections
  - Species
- Economy
  - Trade
  - Travel
  - Product chains
- Politics
  - Voters
  - Relations

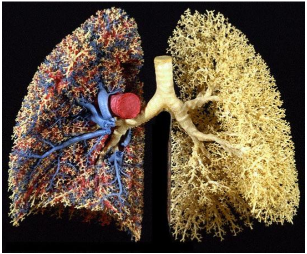
Complexity vs. Complex

Complicated
Torsen differential



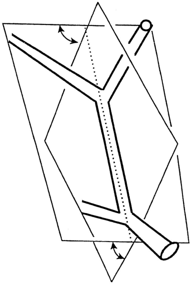


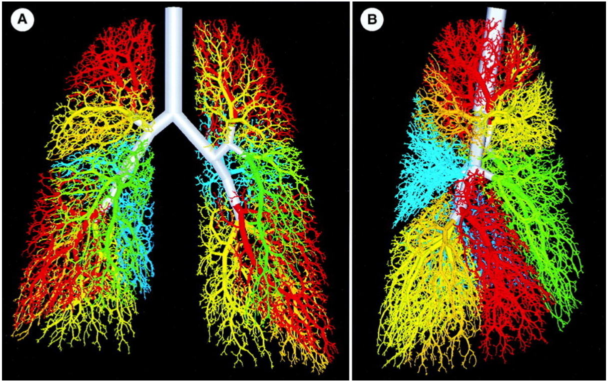
Complex
Bird flock, lungs



Lung

- Angle between child-parent and grandchild-grandparent determines air flow
- Stop if pressure drops below certain value





Complexity

- Complexity, a scientific theory which asserts that some systems display behavioral phenomena that are completely inexplicable by any conventional analysis of the systems' constituent parts. These phenomena, commonly referred to as emergent behaviour, seem to occur in many complex systems involving living organisms, such as a stock market or the human brain.

John L. Casti, Encyclopaedia Britannica

Complexity

- **Many interacting components**
  - Particles: $10^3 - 10^{23}$
  - Brain: $10^3 - 10^{11}$
  - Humans: $34 - 10^9$
  - Computers: $1000 - 10^9$

Complexity

- **Many interacting components**
- **Emergence:** occurs when an entity is observed to have properties its parts do not have on their own
  - *More is different*, P.W. Anderson
  - Brain: neurons $\rightarrow$ thoughts
  - Humans: people $\rightarrow$ society
  - Technology: interconnected computers $\rightarrow$ WWW
  - Particles: crystal structure

Complexity

- **Many interacting components**
- **Emergence**
- **Nonlinearity**
  - Brain: neurons
  - Humans: Reactions
  - Technology: virus spreading
  - Particles: three planet problem

Complexity

- **Many interacting components**
- **Emergence**
- **Nonlinearity**
- **Spontaneous organization**
  - Brain: learning
  - Humans: society
  - Technology: Torrent community
  - Particles: crystals

Complexity

- **Many interacting components**
- **Emergence**
- **Nonlinearity**
- **Spontaneous organization**
- **Diversity**
  - Brain: Different interactions (spontaneous, at will)
  - Humans: society
  - Technology: Torrent community
  - Particles: Phase separation

Networks

- Skeleton of complex systems (units and interactions)
- Underlying network
- Without apprehending this network we cannot understand the complex system $\rightarrow$ Holistic approach

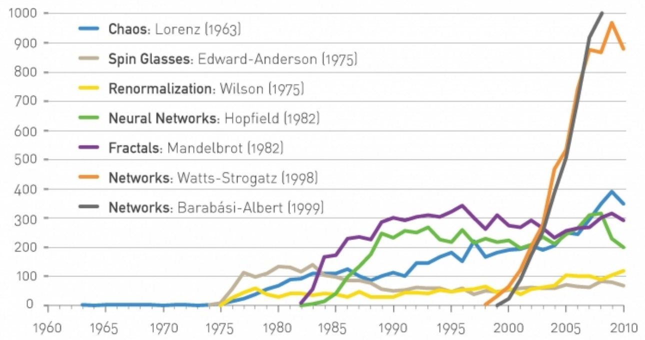
Holism: Looking at systems as a whole is needed for theirs understanding
Reductionism: The precise understanding of the fine details will finally lead to the complete picture

Why now?

- Development of information technology
- Data gathered
- Detailed understanding of building blocks of many systems
- Digitalized world
- Interdisciplinary

Network Science

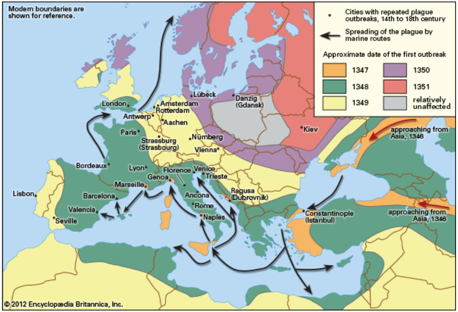
- Citations per year

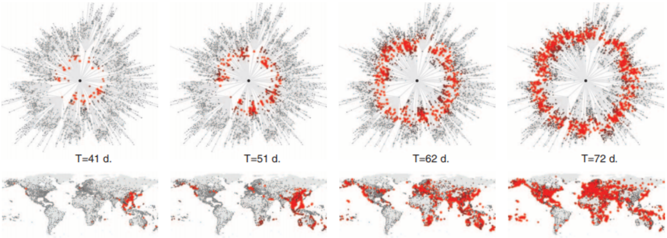


etworksciencebook.com by Barabasi.

What can we learn

- Disease spreading

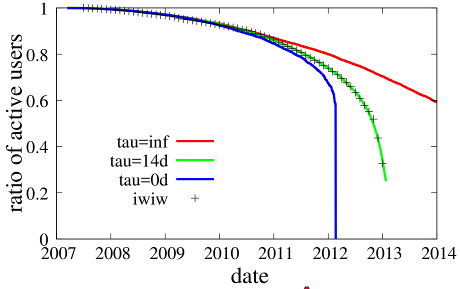


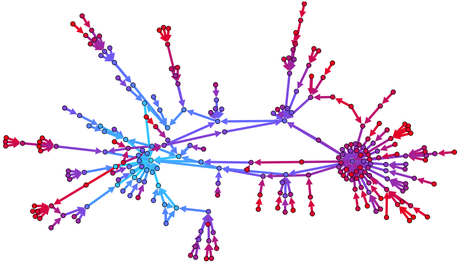


Brockmann-Helbing

What can we learn

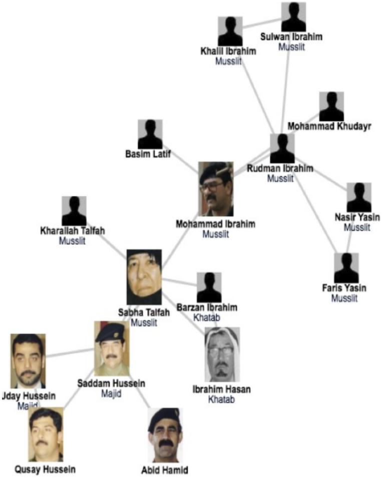
- Disease spreading
- Cascade effects





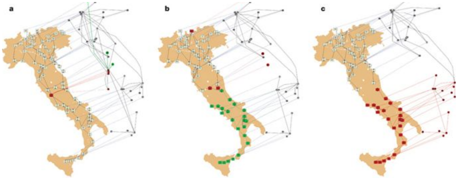
What can we learn

- Disease spreading
- Cascade effects
- Signaling out terrorists



What can we learn

- Disease spreading
- Cascade effects
- Signaling out terrorists
- System robustness

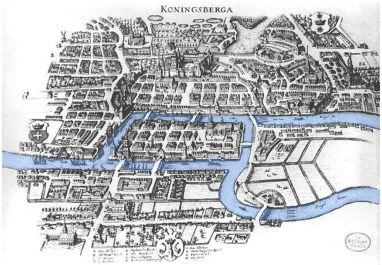


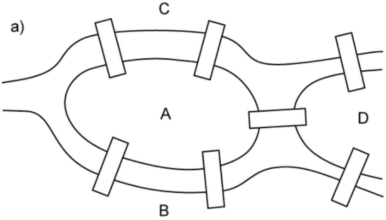
What can we learn

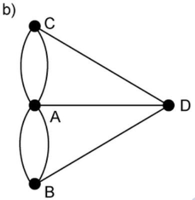
- Disease spreading
- Cascade effects
- Signaling out terrorists
- System robustness
- System efficiency
- Trade efficiency (product suggestions, etc.)

Graph Theory

- Königsberg (Kaliningrad) bridges
- Can we pass all the bridges exactly once?

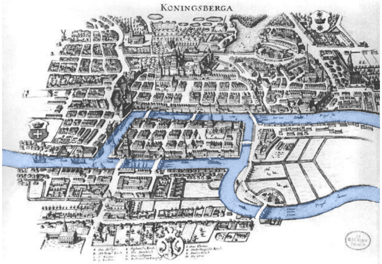






Graph Theory: Euler

- Euler's theorem: An *Eulerian path* on a graph is possible if there are no nodes with odd number of links or there are exactly two such nodes
- A round trip (circle) is possible if there are no nodes with odd number of links.



Wikipedia

Graph Theory: Basics

- Graph:

$$G \equiv \{V, E\}$$

  where

V: vertices (nodes) $(i, j, k, \ldots)$

E: edges (links) $(e_{ij}, \ldots)$

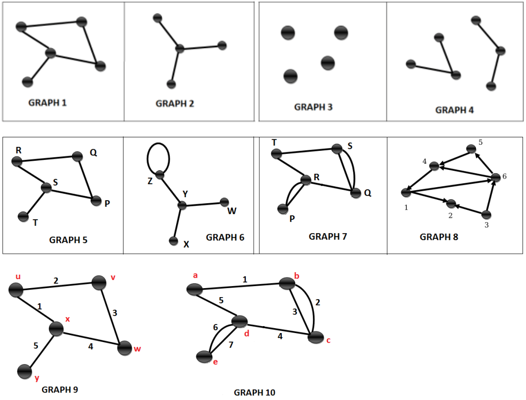
- Network: graph of a system
- Representation:

Nodes: dots

Links: lines between dots

Graph Theory: Types

- Loops: edge starting/ending on the same node: graph 6/Z
- Multiple edges: graph 7/e_{SQ}, e_{PR}
- Directed: graph 8
- Wighted: graph 9,10
- Simple graphs: no loop, no multiple edges, graph 1,2,3,4,5,9
- Bipartite graph: $G = \{U, V, E\}$, $e_{ij} \in E$, $i \in U$, $j \in V$

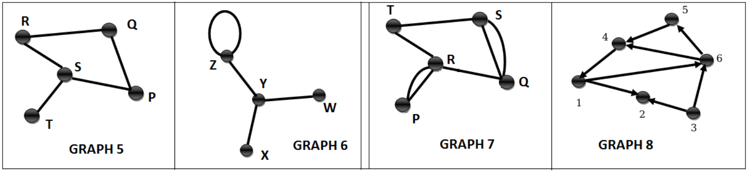


Graph: Adjacency matrix

- Matrix A_{ij} the number of links between nodes i and j

$$A_5 = \begin{pmatrix} & P & Q & R & S & T \\ P & 0 & 1 & 0 & 1 & 0 \\ Q & 1 & 0 & 1 & 0 & 0 \\ R & 0 & 1 & 0 & 1 & 0 \\ S & 1 & 0 & 1 & 0 & 1 \\ T & 0 & 0 & 0 & 1 & 0 \end{pmatrix} \qquad A_6 = \begin{pmatrix} & W & X & Y & Z \\ W & 0 & 0 & 1 & 0 \\ X & 0 & 0 & 1 & 0 \\ Y & 1 & 1 & 0 & 1 \\ Z & 0 & 0 & 1 & 2 \end{pmatrix}$$

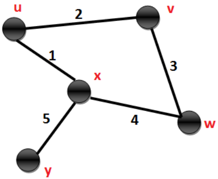
$$A_7 = \begin{pmatrix} & P & Q & R & S & T \\ P & 0 & 0 & 2 & 0 & 0 \\ Q & 0 & 0 & 1 & 2 & 0 \\ R & 2 & 1 & 0 & 0 & 1 \\ S & 0 & 2 & 0 & 0 & 1 \\ T & 0 & 0 & 1 & 1 & 0 \end{pmatrix} \qquad A_8 = \begin{pmatrix} 0 & 1 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 0 \end{pmatrix}$$



Graph Theory: Weighted graphs

- **Weight matrix:**

$$W_9 = \begin{pmatrix} & u & v & w & x & y \\ u & 0 & 2 & 0 & 1 & 0 \\ v & 2 & 0 & 3 & 0 & 0 \\ w & 0 & 3 & 0 & 4 & 0 \\ x & 1 & 0 & 4 & 0 & 5 \\ y & 0 & 0 & 0 & 5 & 0 \end{pmatrix}$$



GRAPH 9

Graph data

- Adjacency matrix: only in memory for small or sparse ones

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

- Edges list: database, data file

```
1 2
1 3
2 3
```

- Adjacency list: generally the most compact, bonus: easy neighbor search

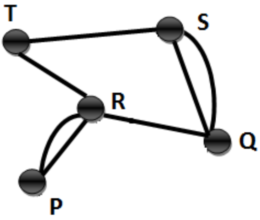
```
1 2 3
2 1 3
3 1 2
```

- Multiline adjacency list: Only for datafile, same but easier to parse

```
1 2
2
3
2 3
1
3
3 2
1
2
```

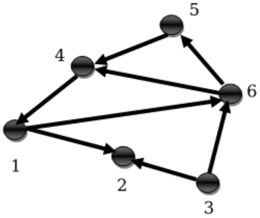
Graph Theory: paths

- **Walk:** sequence of adjacent nodes (connected with edges), e.g. PRQRQST
- **Trail:** sequence of adjacent nodes (connected with edges) where all edges are distinct, e.g. SQSTRPRQ
- **Path:** sequence of adjacent nodes (connected with edges) where all nodes are distinct, e.g. SQRP
- **Circle:** a closed path
- In directed network, the path can follow only the direction of an arrow.



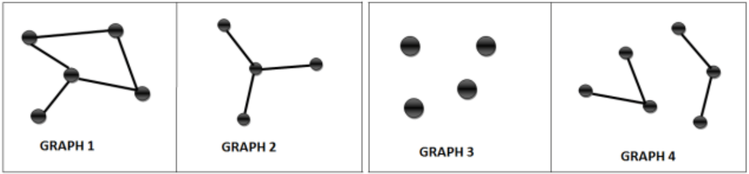
Graph Theory: paths

- **Distance:** The length of the shortest path between two nodes. Length is measured in steps
- There can be more than one shortest paths
- Example:
  - $d(4,6) = 2$
  - $d(6,4) = 1$
  - $d(3,2) = 1$
  - $d(2,3) = \infty$



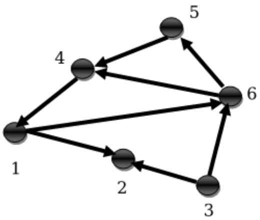
Graph Theory: components

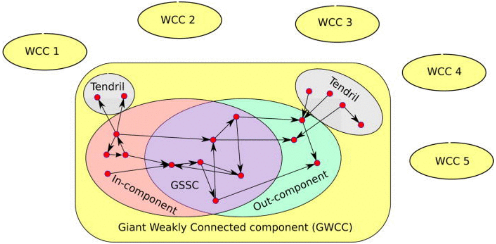
- **Components, clusters:** Set of nodes, with at least one path between any pair of them. (An isolated node is also considered as a component.)
- A graph is connected if it consists of only one component



Graph Theory: components

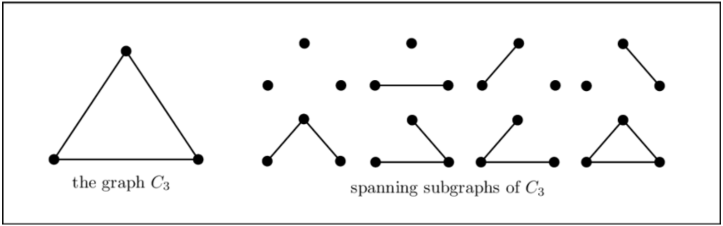
- Component is not trivial for directed graph
- **Strongly connected:** path in both direction between all pair of nodes.
- **Weakly connected:** the undirected version is connected





Graph Theory: subgraphs

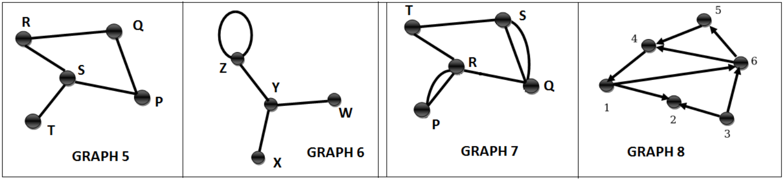
- **Subgraph:** $G' = \{V', E'\}$ is subgraph of $G = \{V, E\}$ if $V' \subseteq V$, $E' \subseteq E$ and all endpoints of E' are in V'
- **Spanning subgraph:** $V' = V$
- **Tree:** A graph where no circles are possible
- **Spanning tree:** A spanning subgraph with no circles



Graph Theory: degree

- **Degree:** The degree of a node is the number of links of a node: $k_i = |\{e_{ij} \in E\}|$
- **In Degree:** In directed graphs: $k_i = |\{e_{ji} \in E\}|$
- **Out Degree:** In directed graphs: $k_i = |\{e_{ij} \in E\}|$
- Example:

Graph 5: k_P=2, k_Q=2, k_R=2, k_S=3, k_T=1
Graph 6: k_W=1, k_X=1, k_Y=3, k_Z=3
Graph 7: k_P=2, k_Q=3, k_R=4, k_S=3, k_T=2
Graph 8: In: k_1=1, k_2=2, k_3=0, k_4=2, k_5=1, k_6=2
Graph 8: Out k_1=2, k_2=0, k_3=2, k_4=1, k_5=1, k_6=2



Graph Theory: degree distribution

- Moments: mean, variance
- Distribution: which function fits it most, where is its maxima, etc.
- **Degree distribution:**
  - $n(k)$, number of nodes with degree k
  - $P(k)$, probability that a node has degree k, $NP(k) \equiv n(k)$
- **Average degree:**

$$\langle k \rangle = \frac{1}{N} \sum_{i=1}^{N} k_i = \sum_{i=1}^{N} kP(k)$$

- For directed graphs, of course, we have $\langle k^{in} \rangle$ and $\langle k^{out} \rangle$

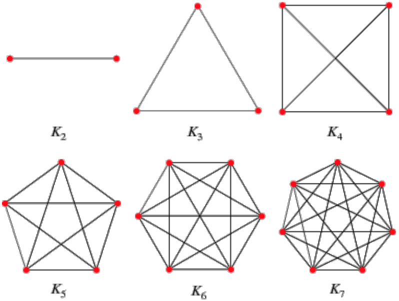
Graph Theory: average degree

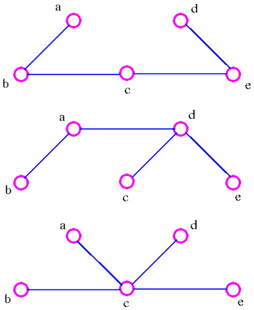
- Complete graphs: $L = N(N-1)/2$, (A link gives two contacts!)

$$\langle k \rangle = \lim_{N \to \infty} (N-1) = \infty$$

- Spanning tree: $L = N - 1$ (circle minus 1 link),

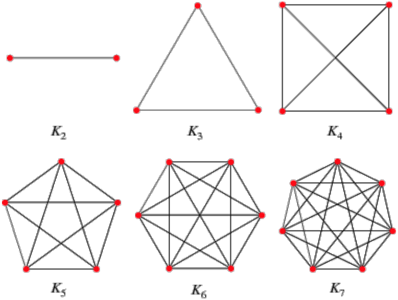
$$\langle k \rangle = \lim_{N \to \infty} 2(N-1)/N = 2$$

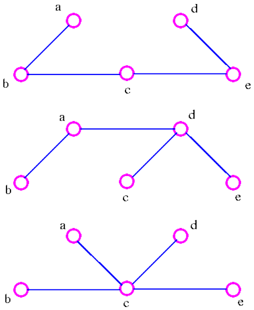




Graph Theory: Sparse-dense graphs

- $L \propto N^{\lambda}$, for large N
  - $\lambda = 1$: Sparse graph
  - $\lambda = 2$: Dense graph
- $\langle k \rangle \propto N^{\mu}$, for large N
  - $\mu = 0$: Sparse graph
  - $\mu = 1$: Dense graph
- Almost all real graphs are sparse





Graph Theory: Adjacency matrix and degree

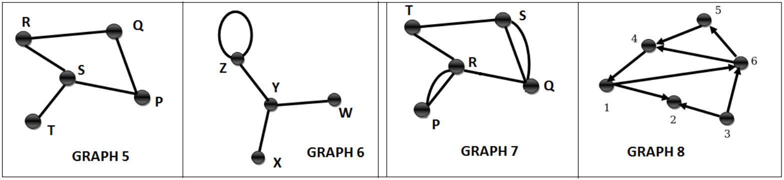
- **Degree:** Undirected, symmetric matrix

$$k_i = \sum_{j=1}^{N} A_{ij} \equiv \sum_{j=1}^{N} A_{ji}$$

- **Degree:** directed, non-symmetric matrix

$$k_i^{in} = \sum_{j=1}^{N} A_{ij}$$

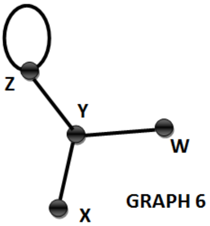
$$k_i^{out} = \sum_{j=1}^{N} A_{ji} \qquad (1)$$



Graph Theory: Powers of the adjacency matrix

- $(A^n)_{ij}$ number of n-step walks between nodes i and j
- **Proof:** Induction. For $n = 1$ trivially true. Assume it is true for $n - 1$. All n-walks to j come from $n - 1$ walks to a neighbor k of j, provided there is a link from k to j.

$$A_6 = \begin{pmatrix} & W & X & Y & Z \\ W & 0 & 0 & 1 & 0 \\ X & 0 & 0 & 1 & 0 \\ Y & 1 & 1 & 0 & 1 \\ Z & 0 & 0 & 1 & 2 \end{pmatrix} \quad A_6^2 = \begin{pmatrix} & W & X & Y & Z \\ W & 1 & 1 & 0 & 1 \\ X & 1 & 1 & 0 & 1 \\ Y & 0 & 0 & 3 & 2 \\ Z & 1 & 1 & 2 & 5 \end{pmatrix}$$



Graph measures: Average distance

- Defined for a single component: average distance between all node pairs:

$$\langle d \rangle = \frac{2}{N(N-1)} \sum_{i \neq j} d_{ij}$$

- Diameter of a network:

$$\delta = \max_{ij} d_{ij}$$

- Usually For large N, $\langle d \rangle \sim \delta \sim N^{\lambda}$
- If $\lambda = 0$ (equiv. logarithmic increase): Small world

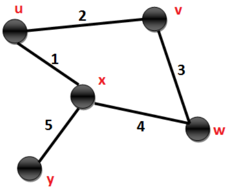
Graph measures: Average distance

- Average distance:

$$\langle d \rangle = \frac{1}{5(5-1)}(\underbrace{1+2+1+2}_{u}+\underbrace{1+1+2+3}_{v}+19) = 1.6$$

- Diameter of a network:

$$\delta = \max_{ij} d_{ij} = 3$$



GRAPH 9

Graph measures: Clustering coefficient

- **Average distance:** fraction of triangles realized out all possible ones at node i

$$C_i = \frac{2n_{\Delta,i}}{k_i(k_i - 1)},$$

where $n_{\Delta,i}$ is the number of triangles at node i.

- Average clustering coefficient:

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

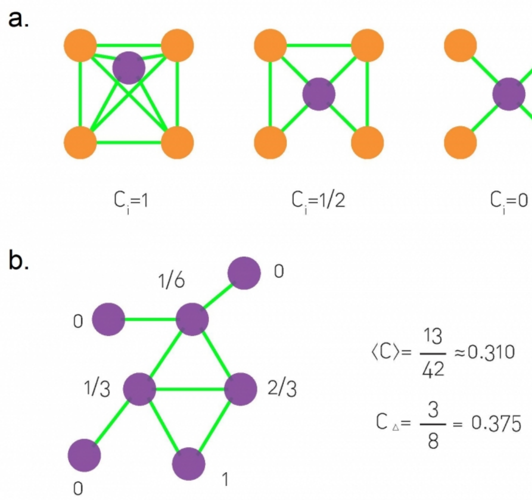
- Global clustering coefficient: fraction of triangles realized out all possible ones.

$$C = \frac{|\{(i,j,k) \text{ circle}, \ i \neq j \neq k\}|}{|\{(i,j,k) \text{ path}, \ i \neq j \neq k\}|}$$

$$C = \frac{3 \times \text{Numberoftriangles}}{\text{Numberofconnectedtriples}}$$

Graph measures: Clustering coefficient

- Example from Barabasi's `http://networksciencebook.com/`

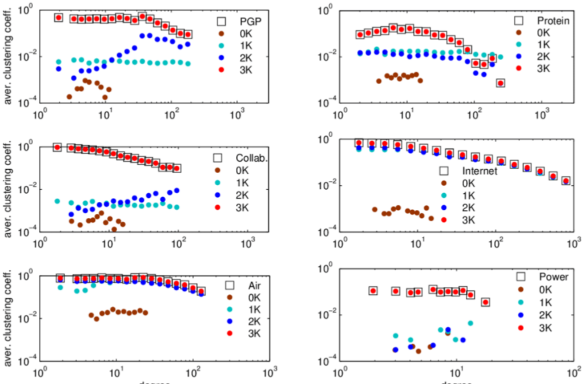


Graph measures: Conditional probability

- **Conditional probability:** $P(x|c)$ normalized distribution of x for cases when condition c holds.
- Example: Clustering coefficient of nodes of degree k:

$$\langle C_k \rangle = \frac{1}{n_k} \sum_{i|k_i=k} C_i(k) = \sum_{C} CP(C|k)$$

https://arxiv.org/pdf/0908.1143



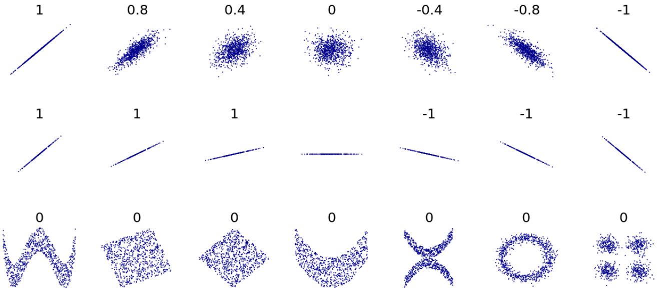
Graph measaures: Pearson correlation coefficient

- **Pearson correlation coefficient:**

$$r_{X,Y} = \frac{E[(X - \mu_X)(Y - \mu_Y)]}{\sigma_X \sigma_Y}$$

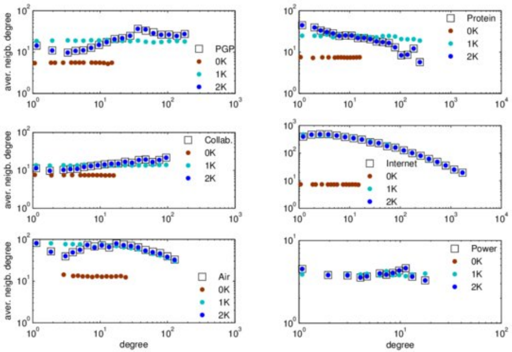
μ_Z mean of Z

σ_Z standard deviation of Z



Graph measaures: Assortativity

- **Assortativity:** $\langle k_{nn}(k) \rangle$ average degree of the neighbors of nodes with degree k
- $\langle k_{nn}(k) \rangle$ increasing $\rightarrow$ assortative mixing
- $\langle k_{nn}(k) \rangle$ decreasing $\rightarrow$ disassortative mixing
- **Assortativity coefficient:** Pearson correlation coefficient of degree between pairs of linked nodes, with $r > 0$ for assortative and $r < 0$ for disassotative mixing.



Graph measures: Assortativity

- **Assortativity:** $\langle k_{nn}(k) \rangle$ average degree of the neighbors of nodes with degree k
- $\langle k_{nn}(k) \rangle$ increasing $\rightarrow$ assortative mixing
- $\langle k_{nn}(k) \rangle$ decreasing $\rightarrow$ disassortative mixing
- **Assortativity coefficient:** Pearson correlation coefficient of degree between pairs of linked nodes, with $r > 0$ for assortative and $r < 0$ for disassotative mixing.

Network	n	r
Physics coauthorship (a)	52 909	0.363
Biology coauthorship (a)	1 520 251	0.127
Mathematics coauthorship (b)	253 339	0.120
Film actor collaborations (c)	449 913	0.208
Company directors (d)	7 673	0.276
Internet (e)	10 697	−0.189
World-Wide Web (f)	269 504	−0.065
Protein interactions (g)	2 115	−0.156
Neural network (h)	307	−0.163
Marine food web (i)	134	−0.247
Freshwater food web (j)	92	−0.276
Random graph (u)		0
Callaway *et al.* (v)		$\delta/(1+2\delta)$
Barabási and Albert (w)		0