Simulations in Statistical Physics

Course for MSc physics students

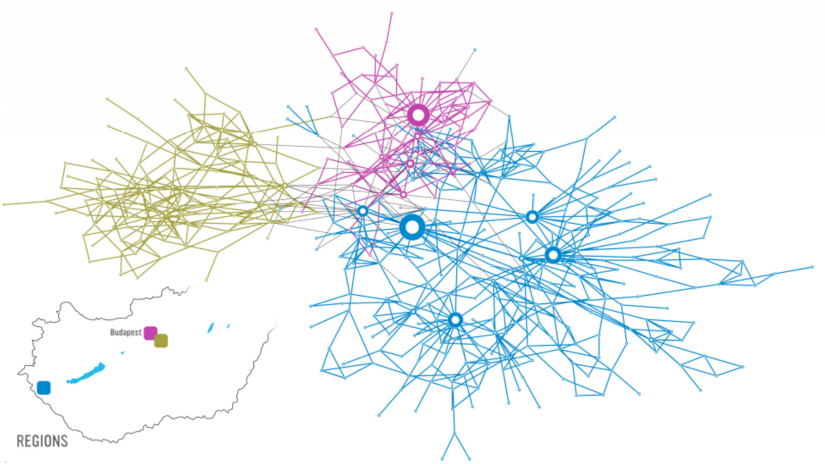
Janos Török

Department of Theoretical Physics

May 4, 2023

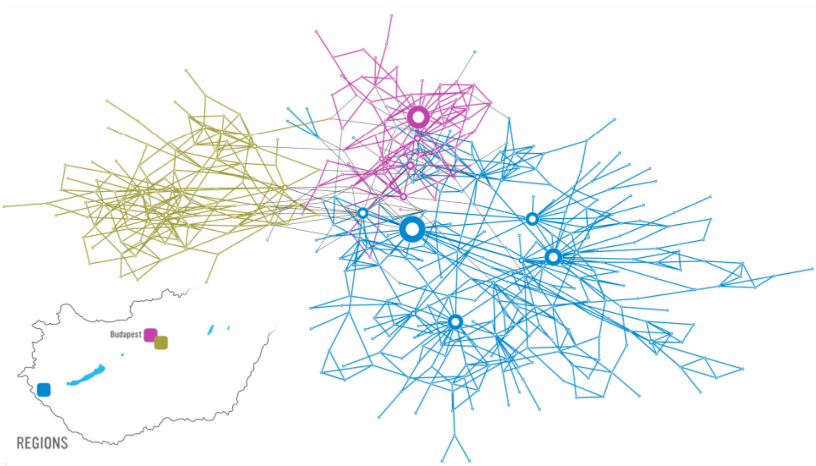
Complex Networks

- Graphs with nontrivial structures
- Graphs consist of nodes and edges connecting nodes



Complex Networks: 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)

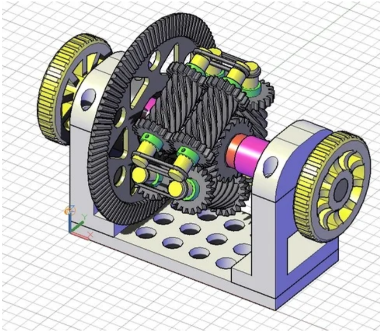
- Biggest hub, and links at distance 1 and 2

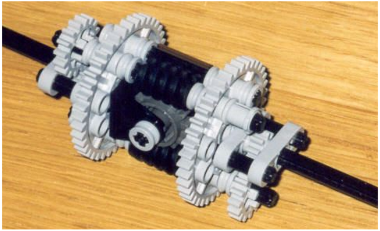
Complex networks

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

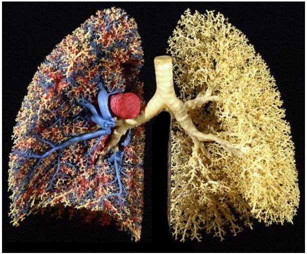
Complexity vs. Complex

Complicated
Torsen differential





Complex
Bird flock, lungs



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

Networks

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

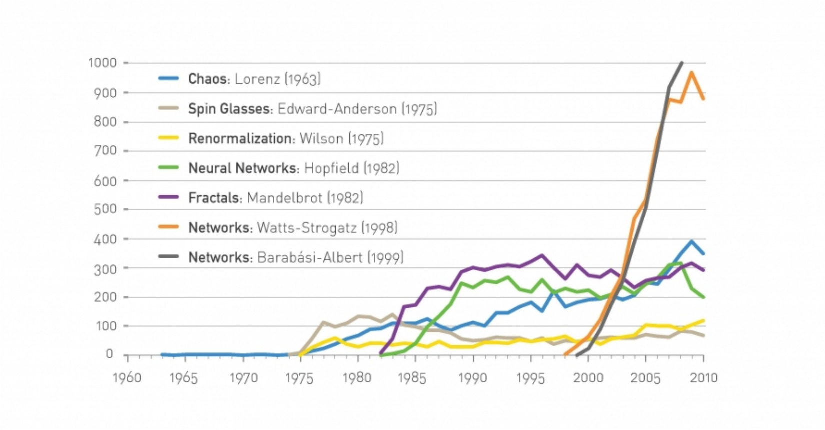
Holism: Looking at systems as a whole is needed for theirs understanding
Reductionism: The precise understanding of the fine details will finally lead to the complete picture

Why now?

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

Network Science

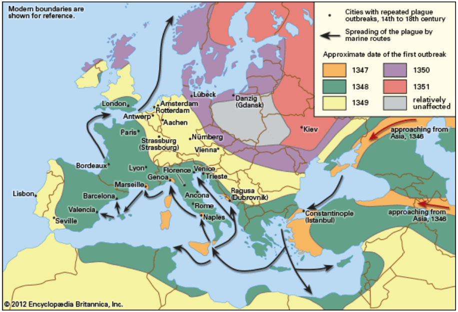
- Citations per year

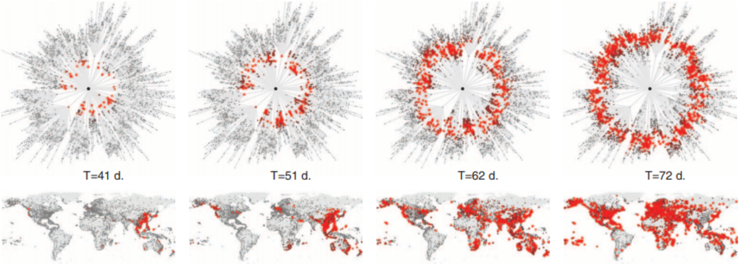


`networksciencebook.com` by Barabasi.

What can we learn

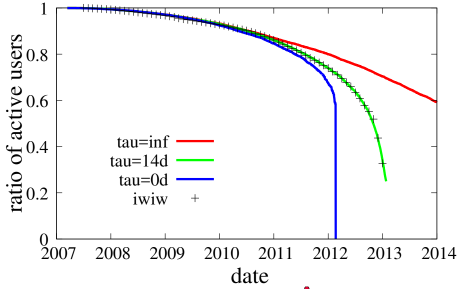
- Disease spreading

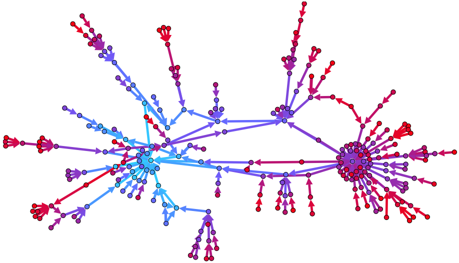




What can we learn

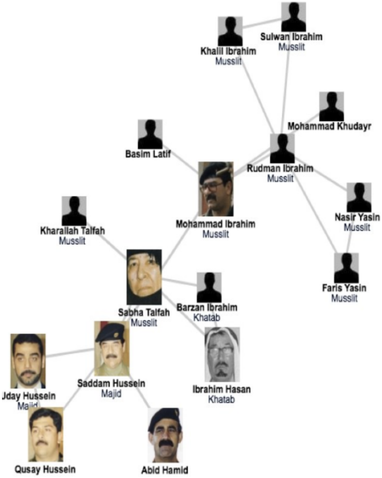
- Disease spreading
- Cascade effects





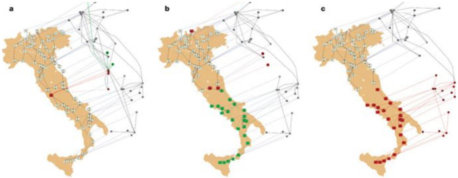
What can we learn

- Disease spreading
- Cascade effects
- Signaling out terrorists



What can we learn

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

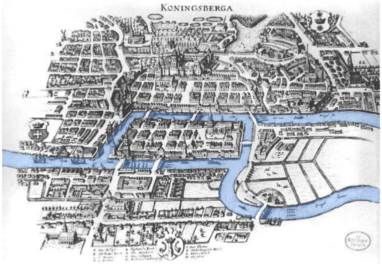


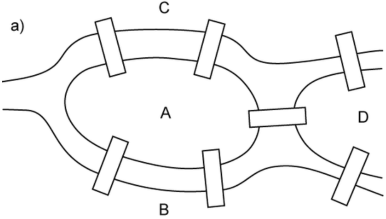
What can we learn

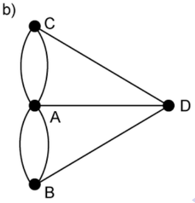
- Disease spreading
- Cascade effects
- Signaling out terrorists
- System robustness
- System efficiency
- Trade efficiency (product suggestions, etc.)

Graph Theory

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

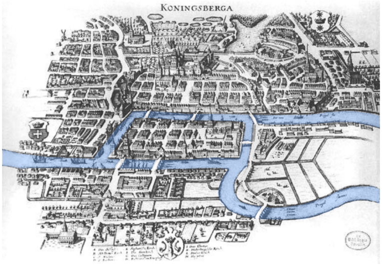






Graph Theory: Euler

- 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) $(i, j, k, \dots)$

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

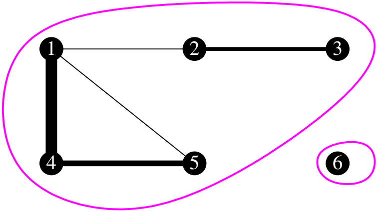
- Network: graph of a system
- Representation:

Nodes: dots

Links: lines between dots

Graphs/Networks

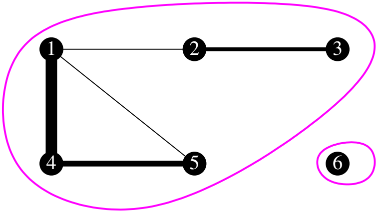
- Described by $\mathcal{G}(V, E)$, where V is the set of vertices, and E is the list of edges
- Alternatively: A_{ij}, Adjacency matrix (1 if there is connection, 0 if not)
- Degree of a node: k number of links connecting to the node (if directed there are *in* k_{ij} and *out* k_{out} degrees)
- A connected component is a subset of the graph in which all vertex pairs are connected by continuous path



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

Adjacency matrix

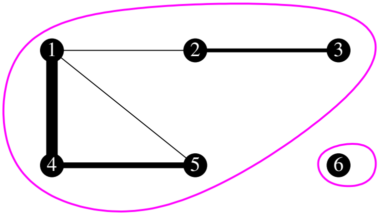
- Adjacency matrix A_{ij}
- 1 if there is connection, 0 if not
- Tells if we can go from node i to node j



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

Adjacency matrix

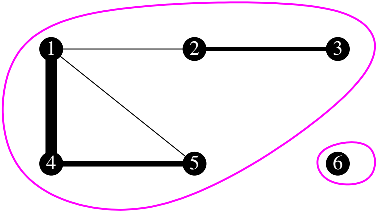
- Adjacency matrix A_{ij}
- 1 if there is connection, 0 if not
- Tells if we can go from node i to node j
- Power n tells how many routes are there from node i to node j



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

Adjacency matrix

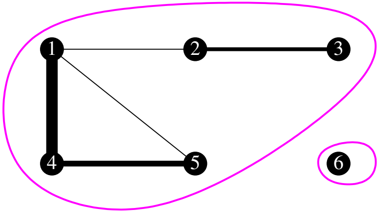
- Adjacency matrix A_{ij}
- 1 if there is connection, 0 if not
- Tells if we can go from node i to node j
- Power n tells how many routes are there from node i to node j



$$A_{ij} = \begin{pmatrix} 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix} \qquad A^3_{ij} = \begin{pmatrix} 2 & 4 & 0 & 4 & 4 & 0 \\ 4 & 0 & 2 & 1 & 1 & 0 \\ 0 & 2 & 0 & 1 & 1 & 0 \\ 4 & 1 & 1 & 2 & 3 & 0 \\ 4 & 1 & 1 & 3 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Weighted graphs

- Described by $\mathcal{G}(V, E)$, where V is the set of vertices, and E is the list of edges
- W_{ij} weight of the link between nodes i and j
- Strength of a node: The sum of weight of the links connecting the node

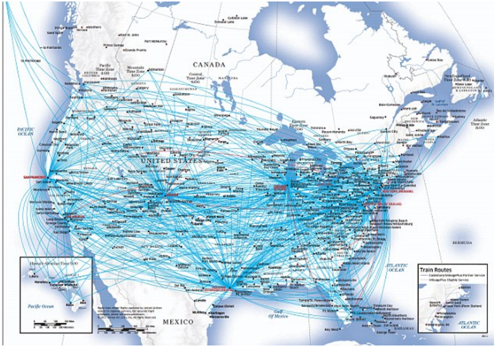


$$W_{ij} = \begin{pmatrix} 0 & 1 & 0 & 7 & 1 & 0 \\ 1 & 0 & 3 & 0 & 0 & 0 \\ 0 & 3 & 0 & 0 & 0 & 0 \\ 7 & 0 & 0 & 0 & 4 & 0 \\ 1 & 0 & 0 & 4 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}$$

Basic network properties

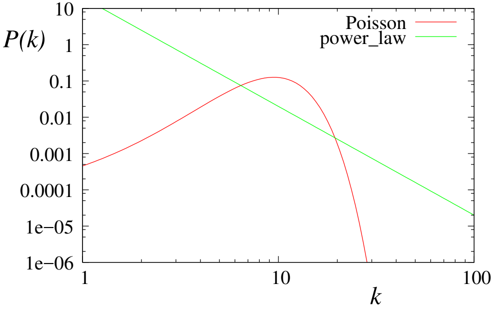
- Global:
  - Degree distribution
  - Shortest path
  - Diameter, small world
  - Clustering coefficient
- Mesoscopic:
  - Communities, modularity
  - Treeness
  - Hierarchy
  - Core-periphery
- Microscopic:
  - Assortativity
  - Centrality

Degree distribution



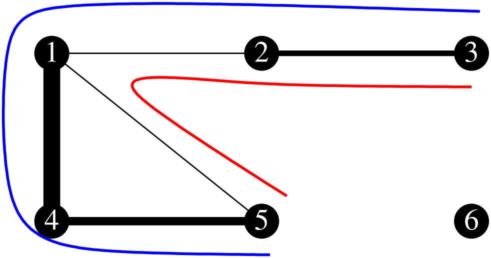
Degree distribution

- Poisson: Well defined mean and variance
- Power law (scale free): Variance and event mean can be undefined, but definitely mode does not match with average
- Existence of the hubs!



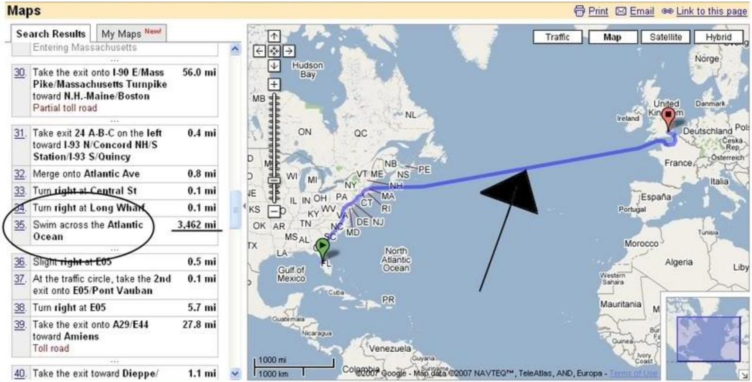
Minimal path

- Minimal path is the path with the smallest possible edges between the two nodes
- If weighted then generally $1/w_{ij}$ is considered (weight is proportional to throughput)
- Many applications: e.g. Route planning



Minimal path

- Minimal path is the path with the smallest possible edges between the two nodes
- If weighted then generally $1/w_{ij}$ is considered (weight is proportional to throughput)
- Many applications: e.g. Route planning



Dijkstra's algorithm

- Find the shortest path from a source
- Known: links, link weights (node distances)
- Store: distance to that point, link to previous element in shortest path
- List of unvisited nodes sorted by distance to origin (set to infinity if unknown)
- Algorithm:
  1. Choose the unvisited node with the smallest distance to the origin
  2. Visit all its unvisited neighbors: if distance is smaller than the current distance to that point, store it and set link to previous element to the current active node
  3. Mark node as finished
  4. If list of unvisited nodes is not empty, go to 1.

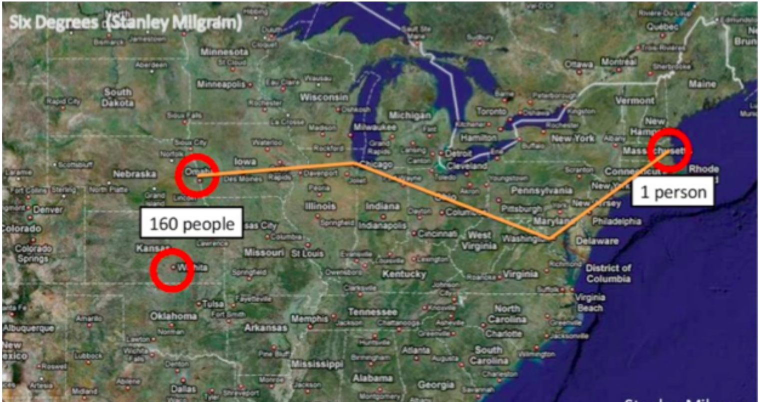
Movie

Diameter/Small world

- Diameter: Largest distance between two vertices
- Average diameter: Mean distance between all vertex pairs
- Society: Small world. Karinthy (1929)
  A fascinating game grew out of this discussion. One of us suggested performing the following experiment to prove that the population of the Earth is closer together now than they have ever been before. We should select any person from the 1.5 billion inhabitants of the Earth – anyone, anywhere at all. He bet us that, using no more than five individuals, one of whom is a personal acquaintance, he could contact the selected individual using nothing except the network of personal acquaintances

Diameter/Small world

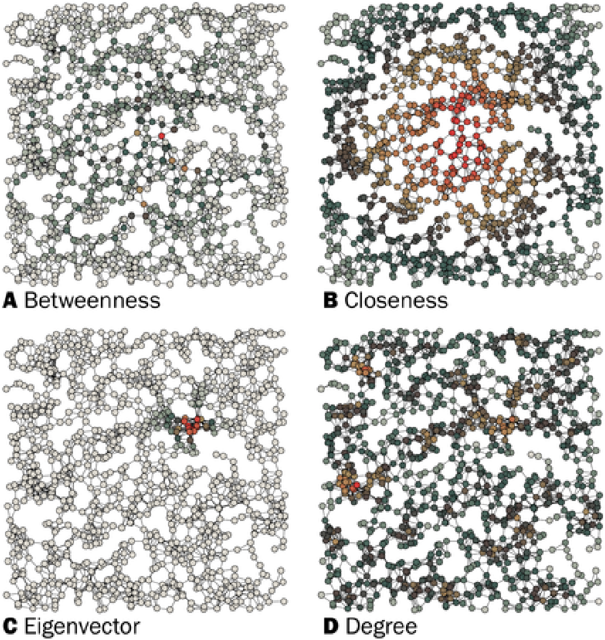
- Diameter: Largest distance between two vertices
- Average diameter: Mean distance between all vertex pairs
- Society: Small world. Karinthy (1929)
- Milgram experiment: Letters were given to individuals in middle us (Kansas/Nebraska)
- They had to reach a person in Boston
- Average hops was 5.5 persons



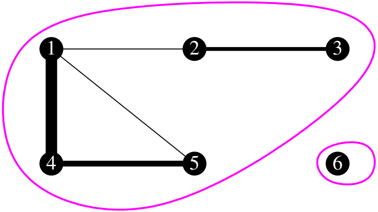
Centrality

- Degree centrality: $C_d(i) = k_i$
- Closeness centrality: inverse of the average distances from i: $C_c(i) = \left(\frac{1}{N-1}\sum_j d_{ij}\right)$
- Betweenness centrality: Number of times a shortest path (σ_{jk} number of shortest paths between j and k) passes through $C_b(i) = \sum_{i \neq j \neq k} \sigma_{jk}(i)/\sigma_{jk}$
- Eigenvector centrality: $Ax = \lambda x$. The eigenvector corresponding to the largest eigenvalue is the centrality measure

Centrality



Centrality



Centrality	1	2	3	4	5	6
Degree	3	2	1	2	2	0
Betweeenness	0.4	0.3	0	0	0	0
Closeness	0.64	0.53	0.35	0.46	0.46	0
Eigenvector	0.6	0.34	0.15	0.50	0.50	0

Random Networks

Generate networks:

- From data:
  - Phone calls
  - WWW links
  - Biology database
  - Air traffic data
  - Trading data
- Generate randomly
  - From regular lattice by random algorithm (e.g. percolation)
  - Erdős-Rényi graph
  - Watts–Strogatz small world model
  - Configuration model
  - Barabási-Albert model

Erdős-Rényi

- P. Erdős, A. Rényi, *On random graphs*, Publicationes Mathematicae Debrecen, Vol. 6 (1959), pp. 290-297 (cit. 789)
- Two variants:
  1. $G(N, M)$: N nodes, M links
  2. $G(N, P)$: N nodes, links with p probability (all considered)
- Algorithm
  1. $G(N, M)$: (If $M \ll N(N-1)/2$)
     - Choose i and j randomly $i, j \in [1, N]$ and $i \neq j$
     - If there is no link between i an j establish one
  2. $G(N, P)$: (Like percolation)
     - Take all $\{i, j\}$ pairs $(i \neq j)$
     - With probability p establish link between i and j

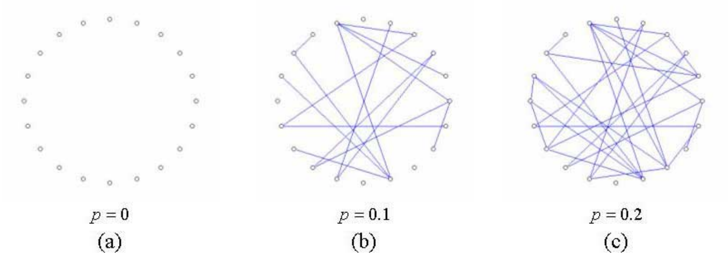
Erdős-Rényi: degree distribution

- Degree distribution

$$P(k) = \binom{N-1}{k} p^k (1-p)^{N-1-k}$$

- For large N and Np =const it is a Poisson distribution

$$P(k) \to \frac{(np)^k e^{-np}}{k!}$$



Erdős-Rényi: Small world/Clustering

- Small world?
  - Yes
  - Average degree $z = 2M/N$
  - Nodes reached after l steps $(z-1)^l$
  - All nodes reached $N = (z-1)^l$ so

$$l = \log N / \log(z-1)$$

  - For humanity: $l \simeq \log(7 \cdot 10^9)/\log(150) = 4.5$
- Clustering
  - Probability of link is independent p
  - Average degree $z = 2M/N$ is kept constant
  - Probability of a link is $p_l = \frac{2M}{N(N-1)}$
  - Clustering

$$C = p_l$$

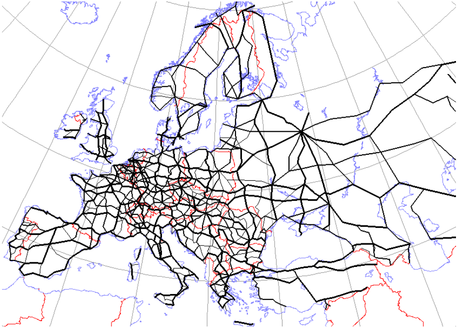
  - For large networks

$$\lim_{N\to\infty} p_l = 0$$

  - In large random networks there are no triangles

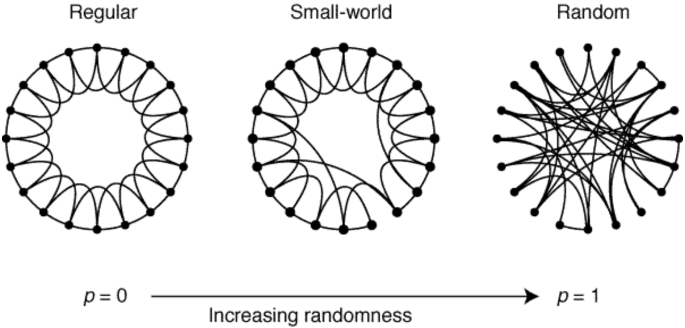
Erdős-Rényi

- Real life: Read networks



Watts-Strogatz model

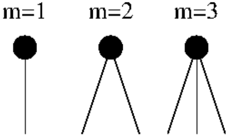
- High clustering: triangular lattice
- Construct a model which continuously extrapolates between the lattice and the random network
- Start from the lattice and randomly rewire links with probability p
- p is a parameter, with $p = 0$ lattice, $p = 1$ Erdős-Rényi



Preferential attachment

Barabási-Albert graph

- Initially a fully connected graph of m_0 nodes
- All new nodes come with m links ($m \leq m_0$)



- Links are attached to existing nodes with probability proportional to its number of links
- k_i is the number links of node i, then

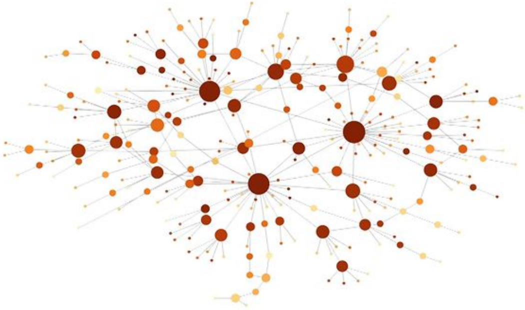
$$p_a = \frac{k_i}{\sum_j k_j}$$

Barabási-Albert graph

- Degree distribution

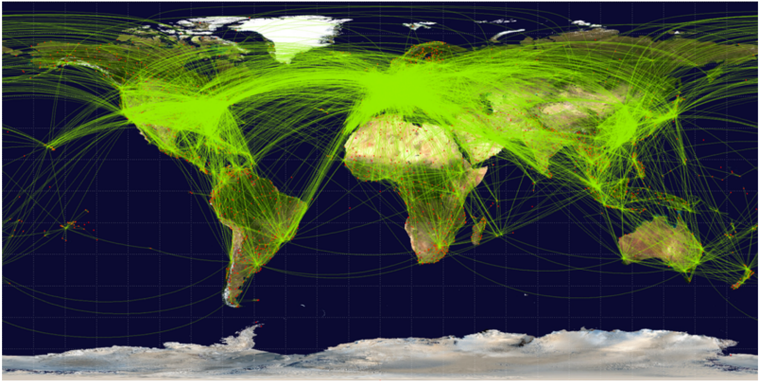
$$p(k) \sim k^{-3}$$

- Independent of m!

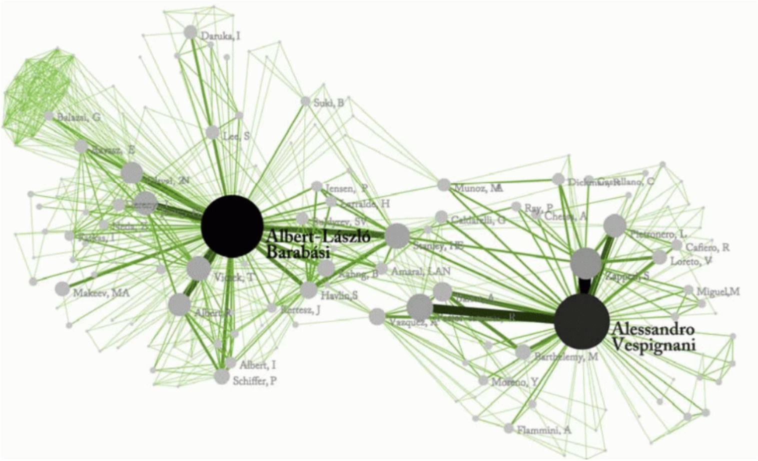


$m = 1$

Scalefree network example: Flight routes



Scalefree network example: Co-authorship



Algorithm for Barabási-Albert graph

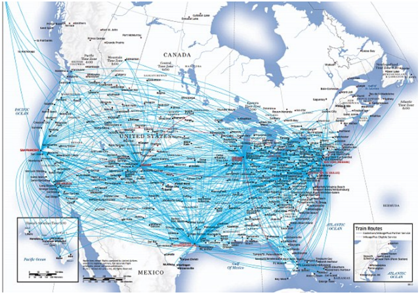
1. $n = m_0$ number of existing nodes
2. $K = \sum_j k_j$ total number of connections
3. r random number $r \in [0, K]$
4. Find $i_{\max}$ for which $\sum_{j=0}^{i_{\max}} k_j < r$
5. If there is no edge then add one between nodes $n + 1$ and $i_{\max}$
6. If node $n + 1$ has less than m connections go to 3.
7. Increase n by 1
8. If $n < N$ go to 2.

Percolation on networks (graphs)

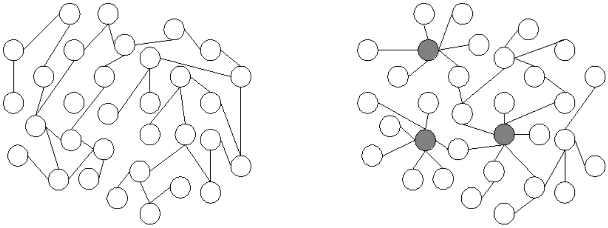
- Network is defined by nodes and links
- Percolation gives us connected components
- Link removal percolation gives information about robustness, and structure

Percolation and attack on random networks

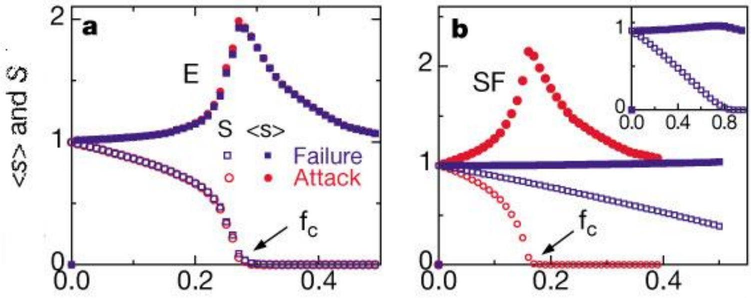
- Failure: equivalent to percolation: remove nodes at random
- Attack: remove most connected nodes



Error vs. attacks

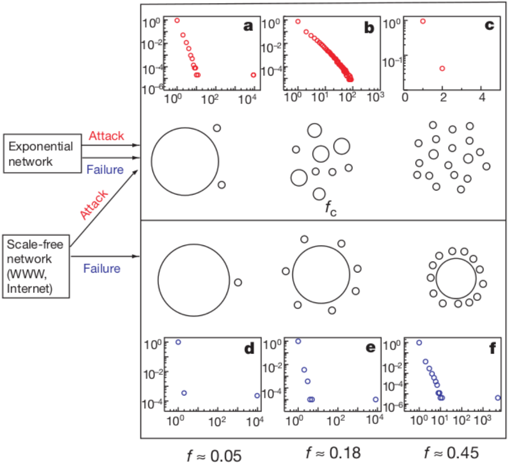


(a) Random network (b) Scale-free network

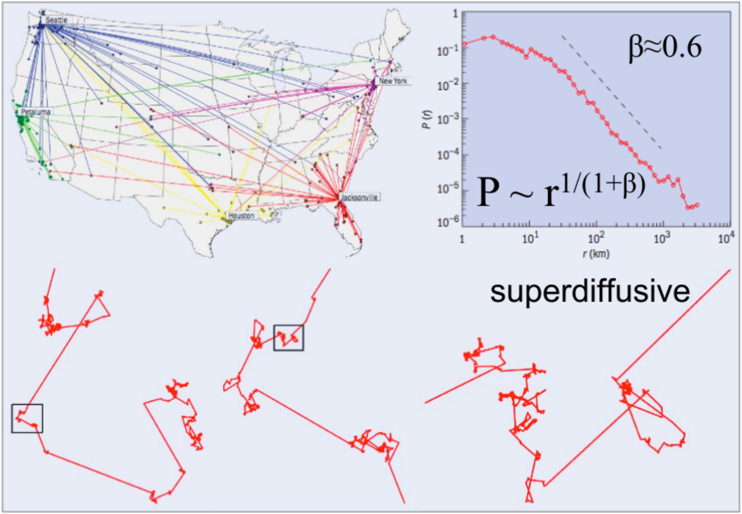


Percolation and attack on random networks

- Failure: equivalent to percolation: remove nodes at random
- Attack: remove most connected nodes



Random Walk on Random Networks



Random Walk on Random Networks

- Rate equation n_k probability of finding the walker an a site with k edges:

$$\frac{\partial n_k}{\partial t} = -rn_k + k \sum_{k'} P(k'|k) \frac{r}{k'} n_{k'}$$

- Uncorrelated random network:

$$P(k'|k) = \frac{k'}{\langle k \rangle} P_{k'}$$

- New equation:

$$\frac{\partial n_k}{\partial t} = -rn_k + r \frac{k}{\langle k \rangle} \sum_{k'} P(k') n_{k'}$$

- Solution:

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

- Random walkers gather on high connectivity nodes

Page rank

- Do what surfers do
- Random walk on pages, but sometimes (probability q) a new (random) restart
- Dumping factor $d = 1 - q$ (general choice $d = 0.85$).
- Self-consistent, equation:

$$P_R(i) = \frac{q}{N} - (1-q)\sum_j A_{ij}\frac{P_R(j)}{k_{\mathrm{out},j}}$$

$$\mathsf{R} = \left(d\mathsf{A} + \frac{1-d}{N}\mathsf{E}\right)\mathsf{R}$$

where E is a matrix of all ones
- Solution: iteration
- Result: Favours sites which are linked by many (reliable sources)

Page rank example