Mathematics of Networks

14.03.2016

Definition of a network

Our definition (Newman): A network (graph) is a collection of vertices (nodes) joined by edges (links).

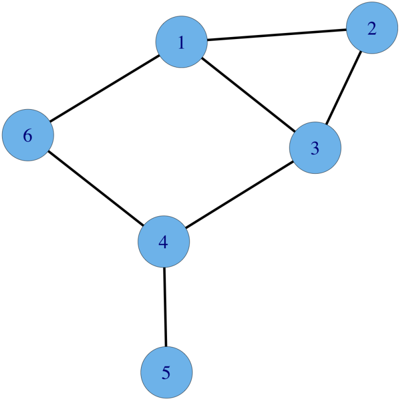
More precise definition (Bollobàs): A graph G is an ordered pair of disjoint sets (V, E) such that E (the edges) is a subset of the set $V^{(2)}$ of unordered pairs of V (the vertices).

Edge list representation

- Vertices:
  1,2,3,4,5,6
- Edges:
  (1,2), (1,3), (1,6),
  (2,3), (3,4), (4,5),
  (4,6)

Definition from graph theory:

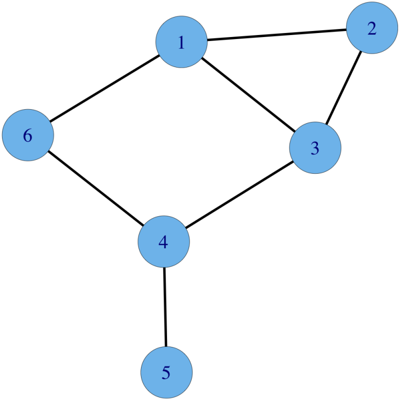
- $n = 6$ vertices: *order* of the graph
- $m = 7$ edges: *size* of the graph



Adjacency matrix **A**

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices i and j} \\ 0 & \text{otherwise.} \end{cases}$$

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

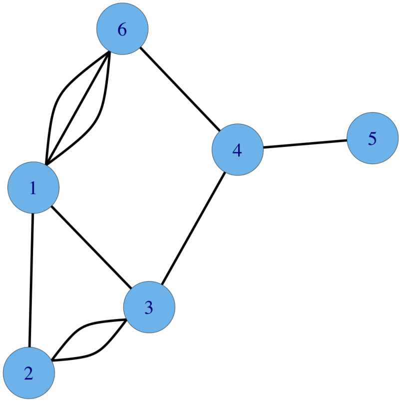


- Diagonal elements are zero.
- Symmetric matrix.

Multigraph

There can be more than one edge between a pair of vertices.

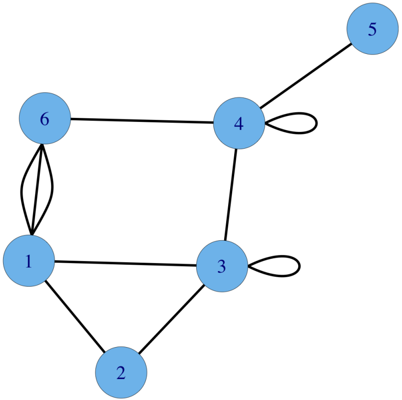
$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 1 & 0 & 0 & 3 \\ 1 & 0 & 2 & 0 & 0 & 0 \\ 1 & 2 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 3 & 0 & 0 & 1 & 0 & 0 \end{pmatrix}$$



Self-edges

There can be self-edges (also called self-loops).

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

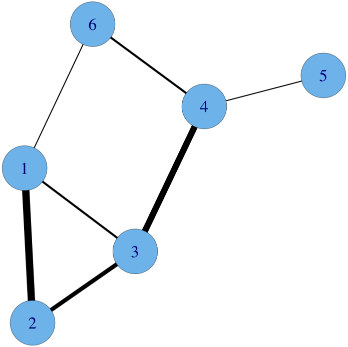


- Diagonal elements can be non-zero:
  Definition: $A_{ii} = 2$ for one self-edge.

Weighted networks

Weight or *strength* assigned to each edge.

$$\mathbf{A} = \begin{pmatrix} 0 & 1.4 & 0.4 & 0 & 0 & 0.8 \\ 1.4 & 0 & 1.2 & 0 & 0 & 0 \\ 0.4 & 1.2 & 0 & 0.2 & 0 & 0 \\ 0 & 0 & 0.2 & 0 & 0.2 & 0 \\ 0 & 0 & 0 & 0.2 & 0 & 0 \\ 0.8 & 0 & 0 & 0.4 & 0 & 0 \end{pmatrix}$$



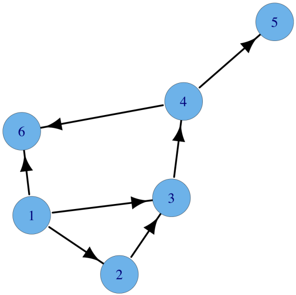
Weights can be both positive or negative.

Directed Networks (Digraphs)

Edge is pointing from one vertex to another (*directed edge*).

$$A_{ij} = \begin{cases} 1 & \text{if there is an edge from } j \text{ to } i \\ 0 & \text{otherwise.} \end{cases}$$

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



In general the adjacency matrix of a directed network is asymetric.

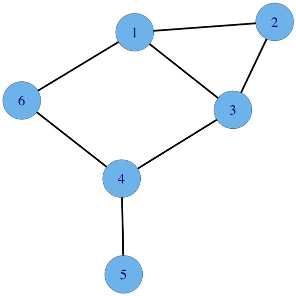
Degree

- Degree k_i of vertex i: Number of edges connected to i.
- Average degree of the network: $\langle k \rangle$.

In terms of the adjacency matrix $\mathbf{A}$:

$$k_i = \sum_{j=1}^{n} A_{ij} \quad , \quad \langle k \rangle = \frac{1}{n} \sum_{i} k_i = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \ .$$

$$k_5 = 1$$
$$k_2 = k_6 = 2$$
$$k_1 = k_3 = k_4 = 3$$
$$\langle k \rangle = 2.33$$



Examples

NETWORK	NODES	LINKS	DIRECTED UNDIRECTED	N	L	⟨k⟩
Internet	Routers	Internet connections	Undirected	192,244	609,066	6.34
WWW	Webpages	Links	Directed	325,729	1,497,134	4.60
Power Grid	Power plants, transformers	Cables	Undirected	4,941	6,594	2.67
Mobile Phone Calls	Subscribers	Calls	Directed	36,595	91,826	2.51
Email	Email addresses	Emails	Directed	57,194	103,731	1.81
Science Collaboration	Scientists	Co-authorship	Undirected	23,133	93,439	8.08
Actor Network	Actors	Co-acting	Undirected	702,388	29,397,908	83.71
Citation Network	Paper	Citations	Directed	449,673	4,689,479	10.43
E. Coli Metabolism	Metabolites	Chemical reactions	Directed	1,039	5,802	5.58
Protein Interactions	Proteins	Binding interactions	Undirected	2,018	2,930	2.90

(from the free textbook "Network Science")

Degree

With n the number of vertices in the graph, and m the number of edges, it holds:

$$2m = \sum_{i=1}^{n} k_i = \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} \ .$$

For the average degree $\langle k \rangle$ of the graph this yields

$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{1}{n} \sum_{i=1}^{n} \sum_{j=1}^{n} A_{ij} = \frac{2m}{n} \ .$$

Density / connectance

Maximum possible number of edges in a simple graph with n vertices:

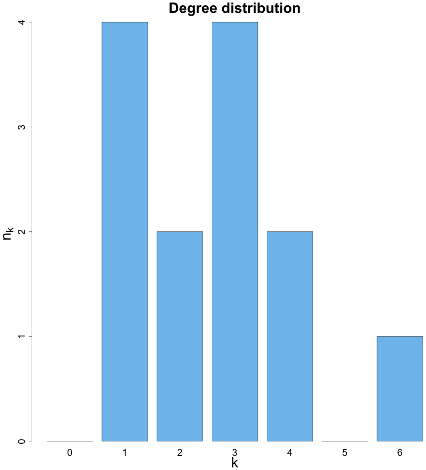
$$\frac{1}{2}n(n-1) \ .$$

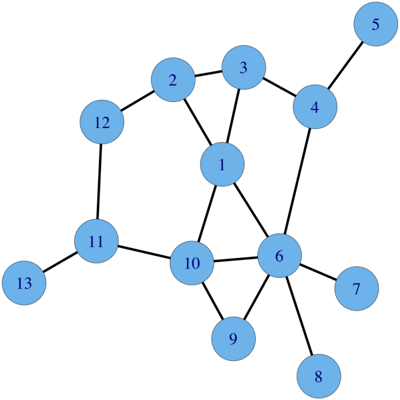
Density or *connectance* of a graph: Fraction of maximum possible number of edges which are present in a given graph:

$$\rho = \frac{m}{\frac{1}{2}n(n-1)} = \frac{2m}{n(n-1)} = \frac{\langle k \rangle}{n-1} \ .$$

Degree distribution

Number of vertices with degree k in a graph: n_k

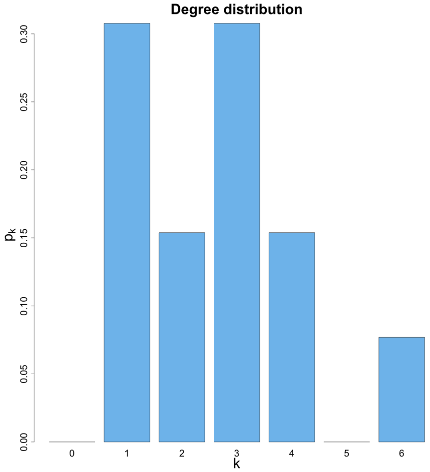


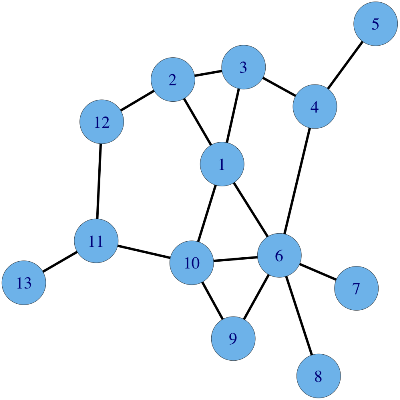


Degree distribution

Fraction of vertices in a graph that have degree k:

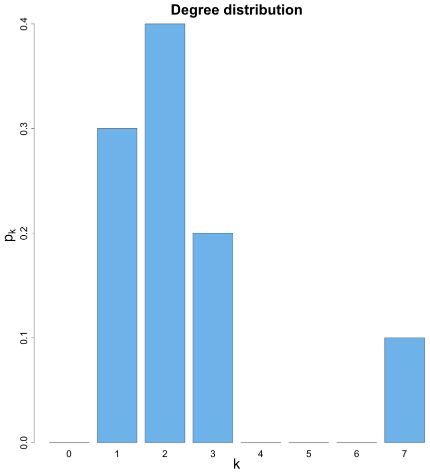
$$p_k = \frac{n_k}{n} \ .$$

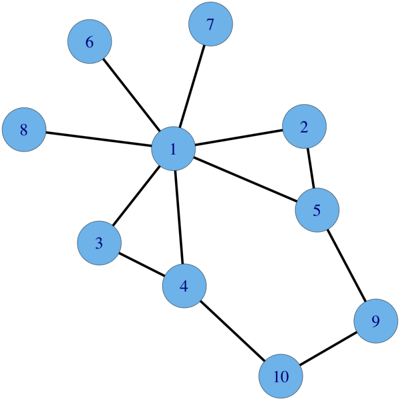




Degree distribution

Hubs: well-connected vertices





Average degree from the degree distribution

Degree distribution tells important information about a network, but doesn't contain the complete information.

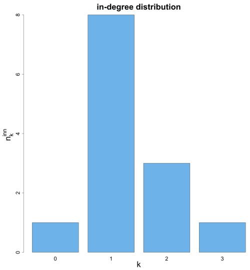
The average degree of a graph can be easily calculated from the degree distribution:

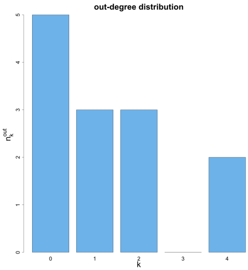
$$\langle k \rangle = \frac{1}{n} \sum_{i=1}^{n} k_i = \frac{1}{n} \sum_{k=0}^{k_{max}} n_k k = \sum_{k=0}^{k_{max}} k p_k \ .$$

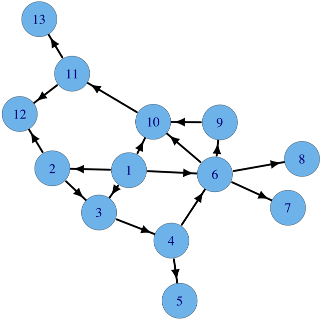
Directed networks: in-degree, out-degree

Number of vertices with k ingoing / outgoing edges.

$$k_i^{in} = \sum_{j=1}^{n} A_{ij} \quad , \quad k_i^{out} = \sum_{j=1}^{n} A_{ji}$$







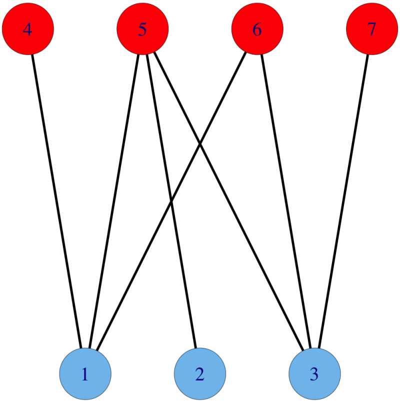
Bipartite networks

Often a system can be represented as a network consisting of two kinds of vertices, with edges only between vertices of different types (group membership). Examples:

- Film actors: Actors, group: Cast of a film
- Coauthorship: Authors, group: Authors of an article
- Rail connections: Train stations, group: Route
- Brazilian soccer players: Players, group: Clubs
- Blinkist: Users, group: Readers of a book

Bipartite networks: Adjacency matrix

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

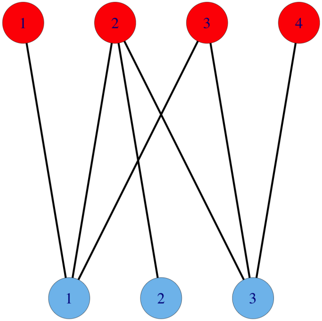


Bipartite networks: Incidence matrix **B**

Vertices of type 1: $i = 1, 2, \ldots, n_1$ (often groups)
Vertices of type 2: $j = 1, 2, \ldots, n_2$ (often people)

$$B_{ij} = \begin{cases} 1 & \text{if there is an edge between vertices i and j} \\ 0 & \text{otherwise.} \end{cases}$$

$$\mathbf{B} = \begin{pmatrix} 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 1 \end{pmatrix}$$



- $n_1 \times n_2$ matrix
- In general asymetric

Bipartite networks: One-mode projections

Projection to a (weighted) network only with vertices of the second type:

$$P_{ij} = \sum_{k=1}^{n_1} B_{ki} B_{kj} \ .$$

That is $\mathbf{P} = \mathbf{B}^T \mathbf{B}$. Note:

$$P_{ii} = \sum_{k=1}^{n_1} B_{ki} B_{ki} = \sum_{k=1}^{n_1} B_{ki} \ .$$

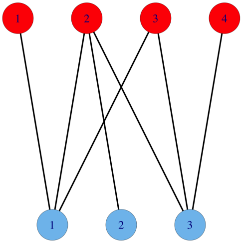
Adjacency matrix ($n_2 \times n_2$):

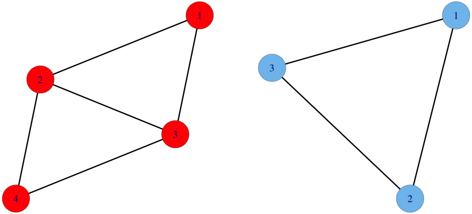
$$A_{ij} = \begin{cases} P_{ij} & \text{if } i \neq j \\ 0 & \text{if } i = j. \end{cases}$$

Bipartite networks: One-mode projections

Two one-mode projections based on

$$\mathbf{P} = \mathbf{B}^T\mathbf{B} \quad , \quad \mathbf{P}' = \mathbf{B}\mathbf{B}^T \ .$$

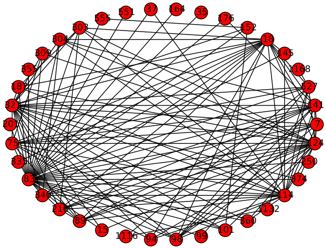


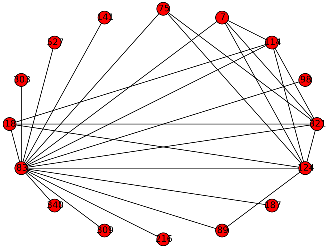


Note: Union of "cliques".

Bipartite networks - Example:Blinkist

One-mode projections to the networks of books, with an edge between two vertices if there more than 1000 / 1500 users have read both books.



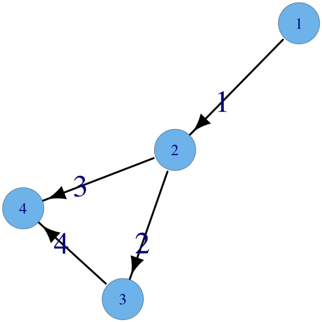


Accessing edge labels with the incidence matrix

For a given network one can consider the edges as one type of vertices of a corresponding bipartite network, with the orginal vertices representing the second type.

Useful for directed networks, where heads and tails of directed edges are represented in the incidence matrix by -1 and 1, respectively.

$$\mathbf{B} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 1 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ 0 & 0 & 1 & -1 \end{pmatrix}$$



Remark: Bipartite networks and cluster synchronization

Sometimes it is interesting to look for a (mostly) bipartite "colouring" of a network.

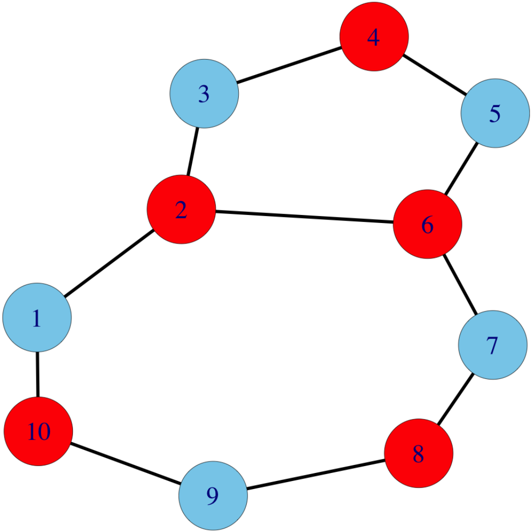
Example: Cluster synchronization of coupled map networks:

$$x_i(t+1) = (1-\epsilon) f[x_i(t)] + \frac{\epsilon}{n} \sum_j f[x_j(t)] \ ,$$

with f a chaotic map, for instance

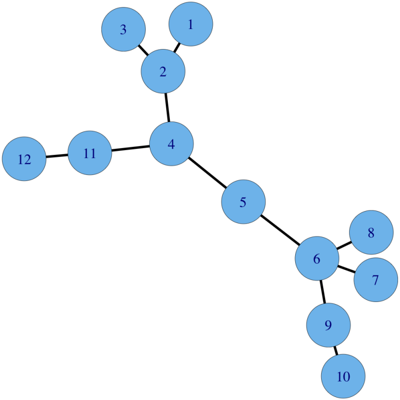
$$f(x) = 2x^2 - 1 \ .$$

Remark: Bipartite networks and cluster synchronization



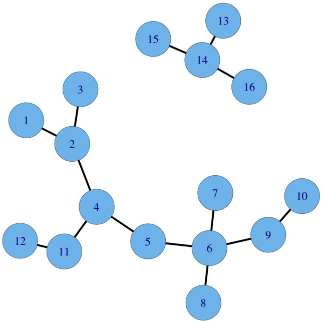
Trees

A *tree* is a connected, undirected network that contains no closed loops.



Trees

- A collection of trees is called a *forest*.
- Trees play an import role for random graph models.
- In a tree, there is exactly one path between any pair of vertices.
- A tree of n vertices always has exactly $n - 1$ edges.
- Any connected network with n vertices and $n - 1$ edges is a tree.

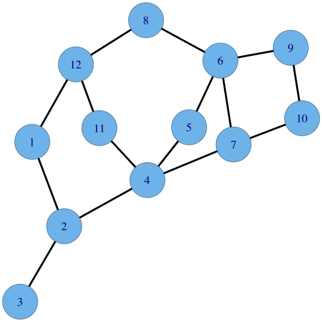


Planar networks

A *planar network* is a network that can be drawn on a plane without having any edges cross.

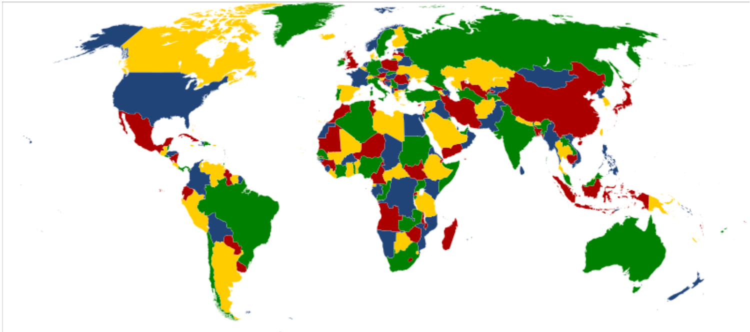
Examples:

- Trees
- Road networks (approximately)
- Power grids (approximately)
- Shared borders between countries, etc.



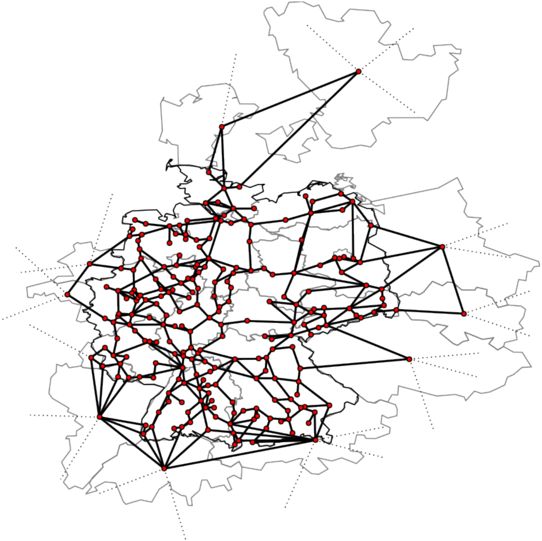
Planar networks - Four-color theorem

"In mathematics, the four color theorem, or the four color map theorem, states that, given any separation of a plane into contiguous regions, producing a figure called a map, no more than four colors are required to color the regions of the map so that no two adjacent regions have the same color." [Wikipedia]



Power Grid expansion optimisation

Expansion condition: planar graph



Paths

- Route through the network, from vertex to vertex along the edges
- Defined for both directed and undirected networks
- Special case: self-avoiding paths
- *Length* of a path: number of edges along the path ("hops")
- Number of paths of length r between vertices i and j:

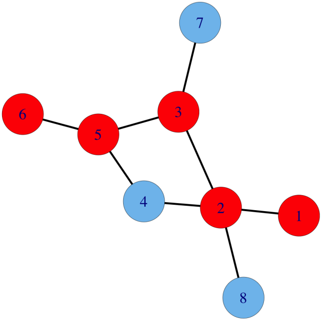
$$N_{ij}^{(r)} = [\mathbf{A}^r]_{ij}$$

- Total number L_r of loops of length r anywhere in the network:

$$L_r = \sum_{i=1}^{n} [\mathbf{A}^r]_{ii} = \mathrm{Tr}\mathbf{A}^r \ .$$

Geodesic / shortest paths

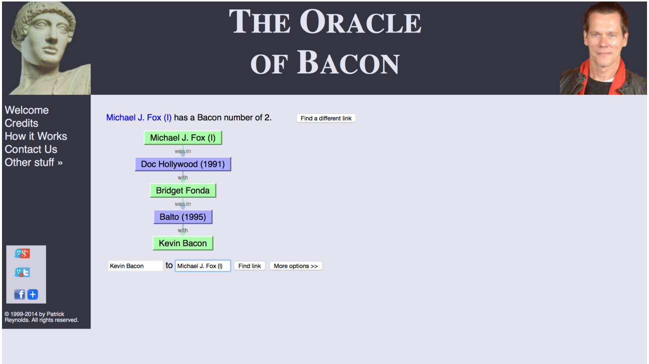
- A path between two vertices such that no shorter path exists
- Geodesic distance between vertices i and j is the smallest value of r such that $[\mathbf{A}^r]_{ij} > 0$.
- Self-avoiding
- In general not unique
- *Diameter* of a network: Length of the longest geodesic path between any pair of vertices



Shortest paths – some examples

Oracle of Bacon: https://oracleofbacon.org/

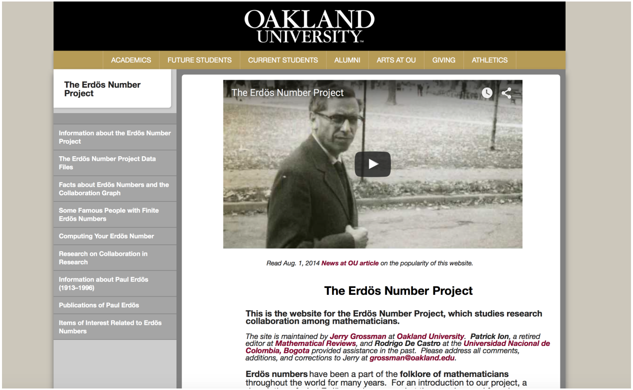
- Network of movie actors (joint appearance in a movie, based on IMDB)
- Geodesic distance to Kevin Bacon



Shortest paths – some examples

Erdös number: Consult http://wwwp.oakland.edu/enp/

- Coauthorship network
- Geodesic distance to Paul Erdös



Shortest paths – "Six degrees of separation"

- Classic experiment by Stanley Milgram (also known for "obedience to authority")
- Average path lengths in social networks

An Experimental Study of the Small World Problem*

JEFFREY TRAVERS
Harvard University
AND
STANLEY MILGRAM
The City University of New York

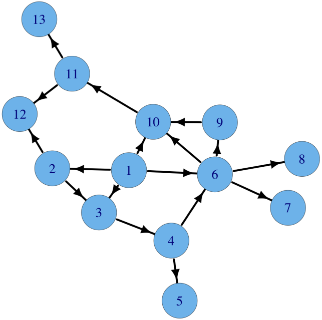
Arbitrarily selected individuals (N=296) in Nebraska and Boston are asked to generate acquaintance chains to a target person in Massachusetts, employing "the small world method" (Milgram, 1967). Sixty-four chains reach the target person. Within this group the mean number of intermediaries between starters and targets is 5.2. Boston starting chains reach the target person with fewer intermediaries than those starting in Nebraska; subpopulations in the Nebraska group do not differ among themselves. The funneling of chains through sociometric "stars" is noted, with 48 per cent of the chains passing through three persons before reaching the target. Applications of the method to studies of large scale social structure are discussed.

Shortest paths and breadth-first search

- Single run of the algorithm: Finds shortest (geodesic) distance from a source vertex s to *every other* vertex in the same component of the network
- In a second step the algorithm also finds shortest paths by construction the so-called *shortest path tree*

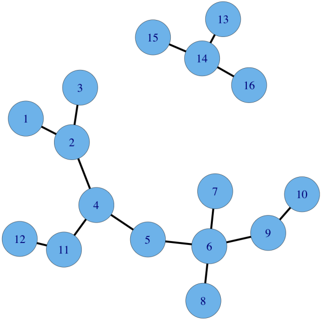
Acyclic directed network

- Directed network without closed loops of edges (DAG)
- Examples: power flow in an electricity grid, citation network of papers
- Topological ordering: For every directed edge $i \to j$, vertex i comes before j in the ordering: (1,2,3,4,6,9,10,11,12,8,7,5,13)
- With a topological ordering, the adjacency matrix of an acyclic directed network is *strictly triangular*



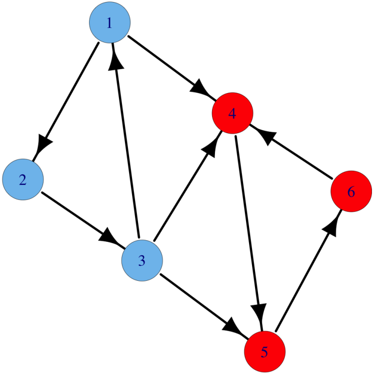
Components of networks

- Subgroups of vertices with no connections between the respective groups
- *Disconnected* network
- Subgroups: *components*
- Adjacency matrix: Block-diagonal form



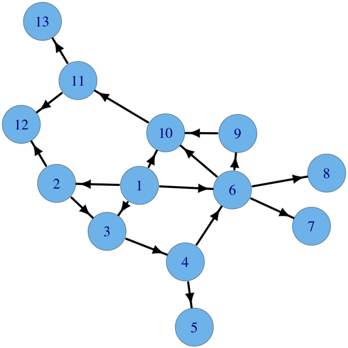
Components in directed networks

- *Weakly connected components*: connected in the sense of an undirected network
- *Strongly connected components*: directed path in both directions between every pair in the subset



Components in directed networks

- *Out-component* of a vertex i: set of vertices which are reachable via directed paths starting form i, including the vertex i itself
- *In-component* of a vertex i: set of vertices from which there is a directed path to i, including the vertex i itself
- One often considers the out- or in-component of a strongly connected component



Network of Global Corporate Control

Ownership network of transnational corporations (TNCs)
Vitali et al., PLOS One, 6 (2011)

- Ownership matrix **W**:
  W_{ij} is the percentage of ownership that the owner (shareholder) i holds in firm j
- If $W_{ij} > 0$ and $W_{jl} > 0$, then vertex i has an *indirect* ownership of firm l
- Data: Orbis 2007 database
- Resulting network: 600508 vertices (economics actors), containing 43060 TNCs, 1006987 edges (ownership ties)

Network of Global Corporate Control

Vitali et al., PLOS One, 6 (2011)