COMPLEX NETWORKS

János Kertész

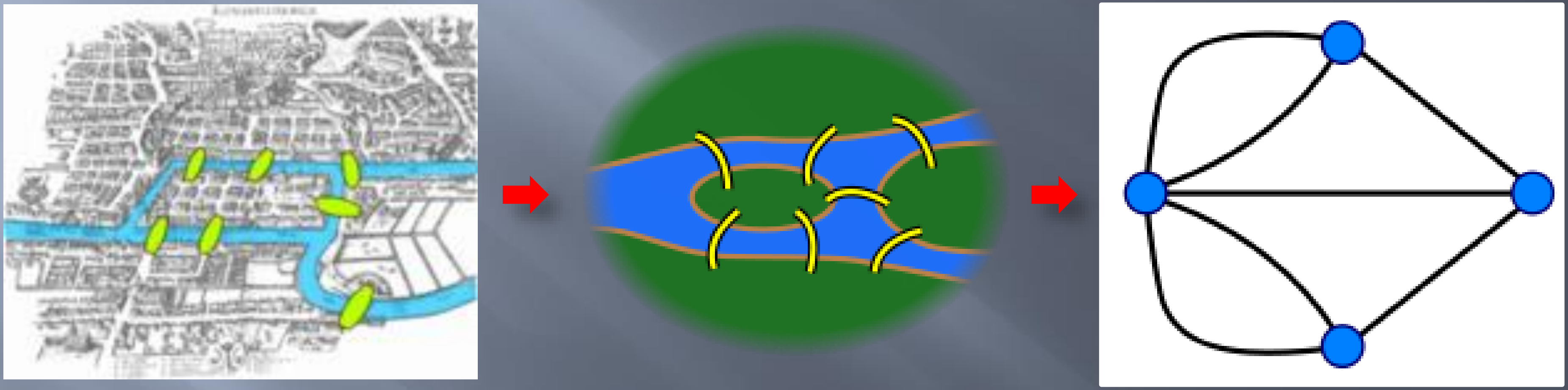
janos.kertesz@gmail.com

3. BASIC NOITIONS OF NETWORK CHARACTERIZATION

Graph theory: history

The problem of Königsberg bridges

Königsberg and the river Pregel

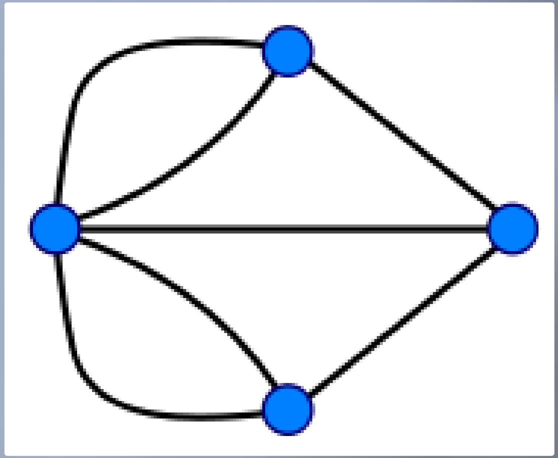


Question: Is it possible to traverse all bridges exactly once in a walk? Is it possible to make such a round trip?

Steps of abstraction: a graph is useful if connectedness, topology of interactions are asked for.

Wikiepdia

Graph theory: history



Is it possible to draw this line without lifting the pencil?

Leonhard Euler (1735): No!

Euler

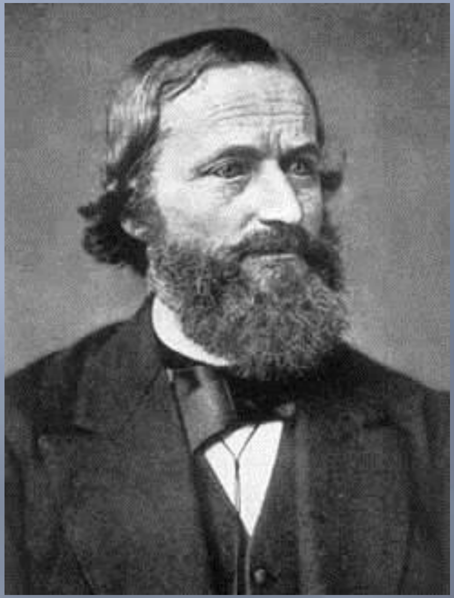
Euler's theorem: An "Eulerian path" on a graph is possible if there are no nodes with odd number of links or there are exactly two such nodes. A round trip (cycle) is possible if there are no nodes with odd number of links.

(Not to be confused with Hamiltonian paths and cycles)

Wikiepdia

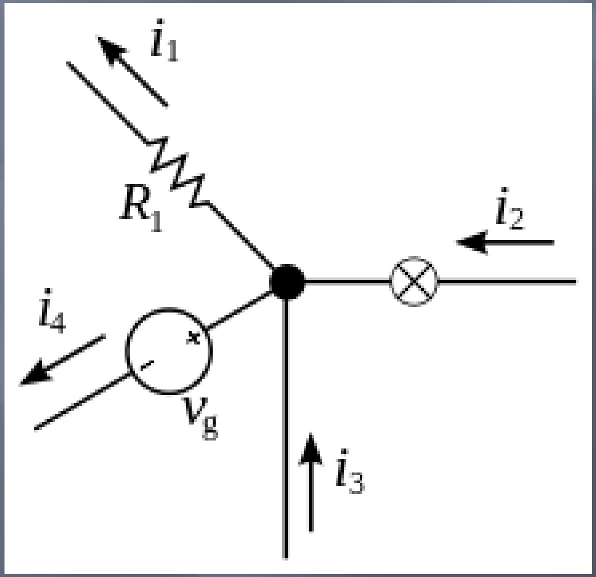
Graph theory: history

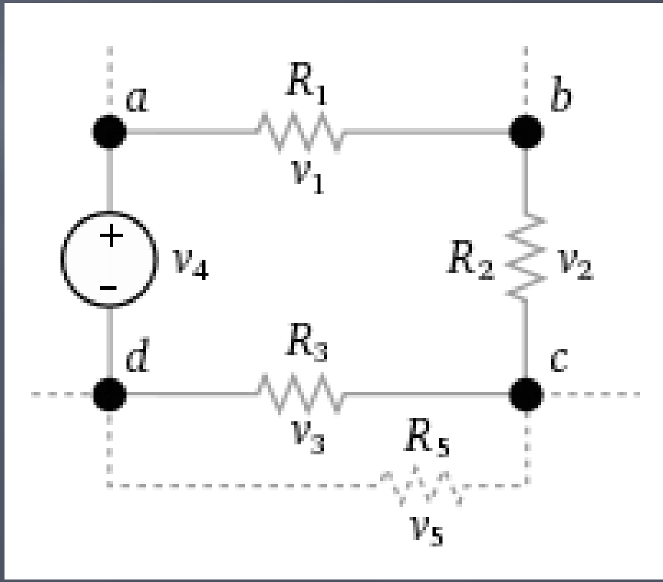
Kirchhoff’s two laws of electrical circuits (1845)



G. Kirchhoff

1. Sum of currents at a node is 0
2. Sum of voltages along a circle is 0

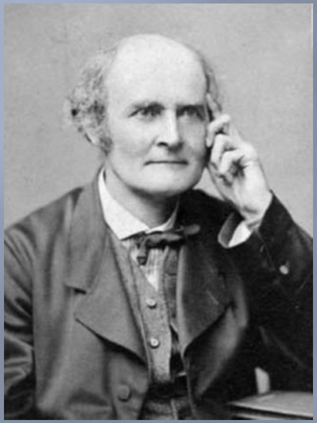


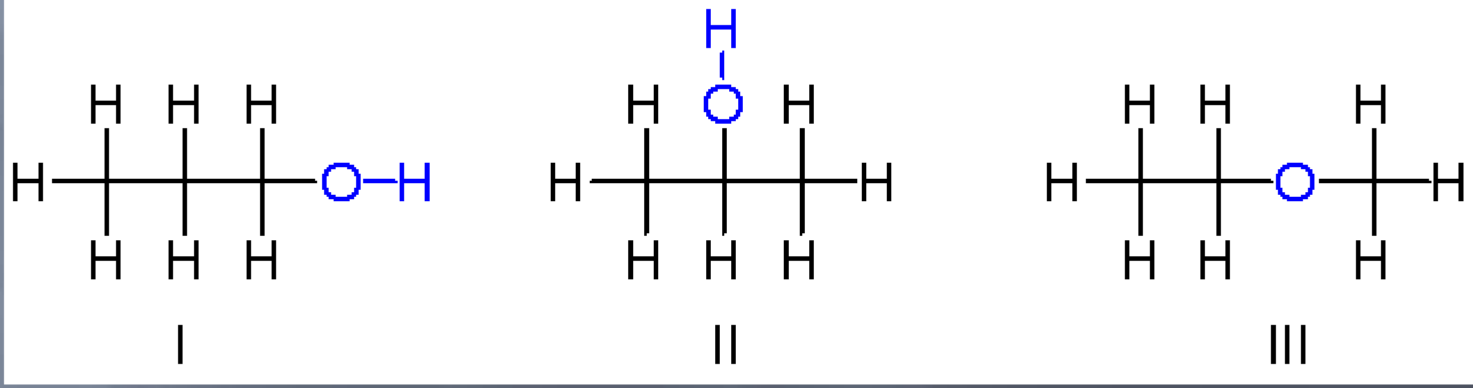


Wikiepdia

Graph theory: history

Enumeration of chemical isomers:
How many ways can atoms be connected if their valence (and possibly binding preferences) are given?





Arthur Cayley
1887

György Pólya:

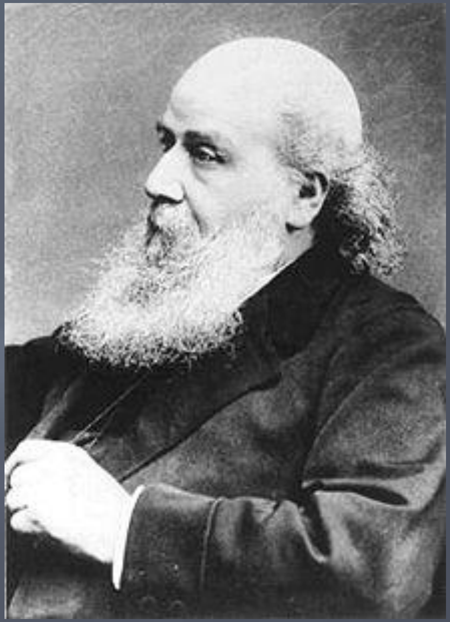
Graph theory in Chemistry (1935)

Wikiepdia

Graph theory: history

The term “graph” was coined by James Joseph Sylvester (1878)

Graph theory has been used in:
Chemistry,
Electrical engineering,
Traffic planning
Social sciences
and many more fields



First textbook: Dénes König (1936)

Wikiepdia

Graph theory: basics

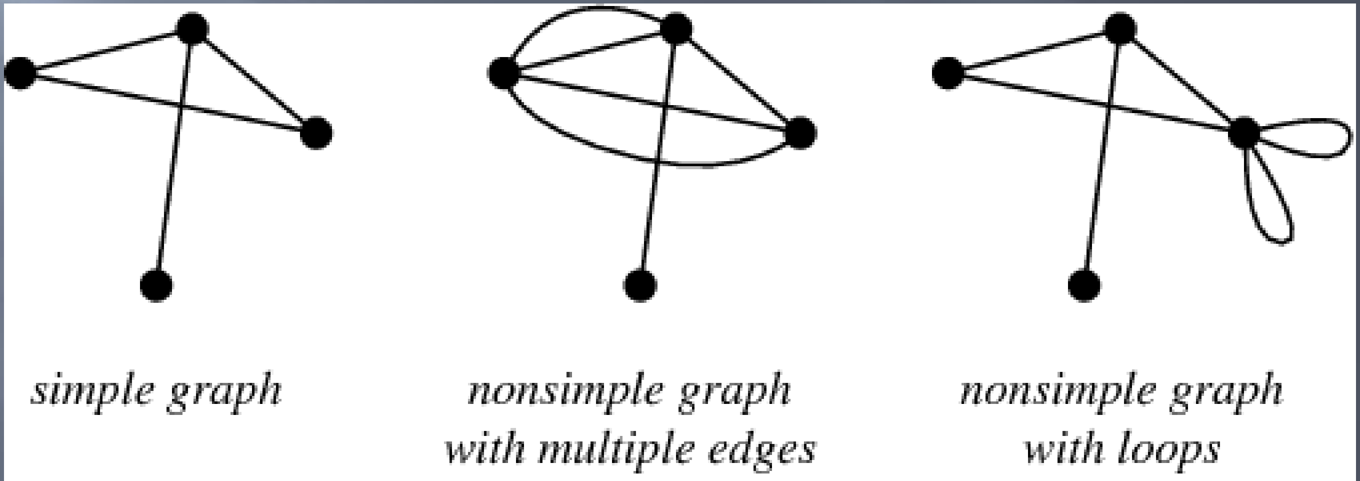
Graph: $G \equiv \{V, E\}$

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

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

Network is the graph of a system.

G can be represented by drawing nodes as dots and links as lines connecting them.



simple graph

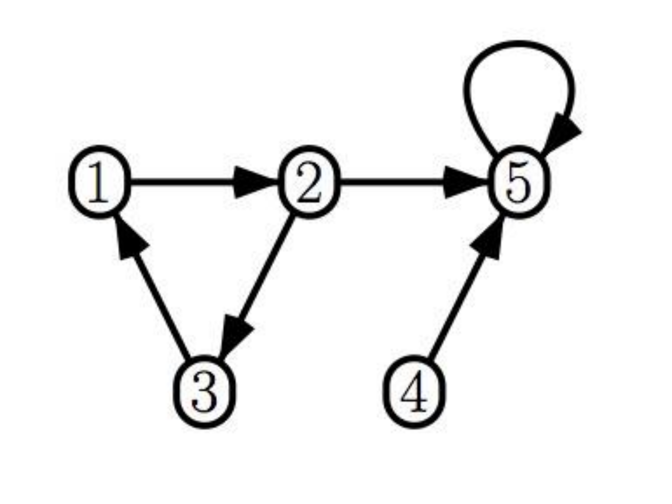
nonsimple graph with multiple edges

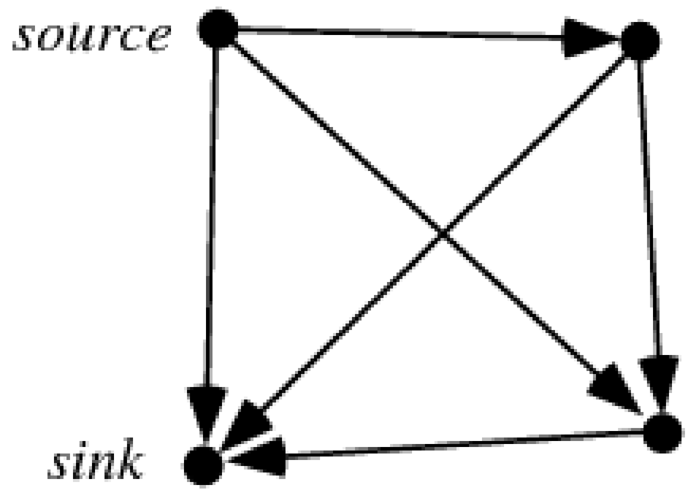
nonsimple graph with loops

http://mathworld.wolfram.com/SimpleGraph.html

Graph theory: basics

Directed graph: In elements of the set *E* the order of the nodes matter: $e_{ij} \neq e_{ji}$. The directed edges are represented by arcs.





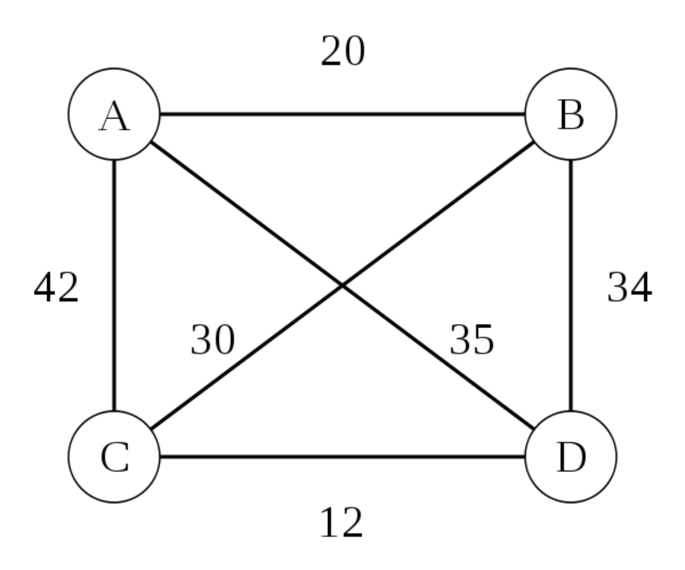
Wikimedia

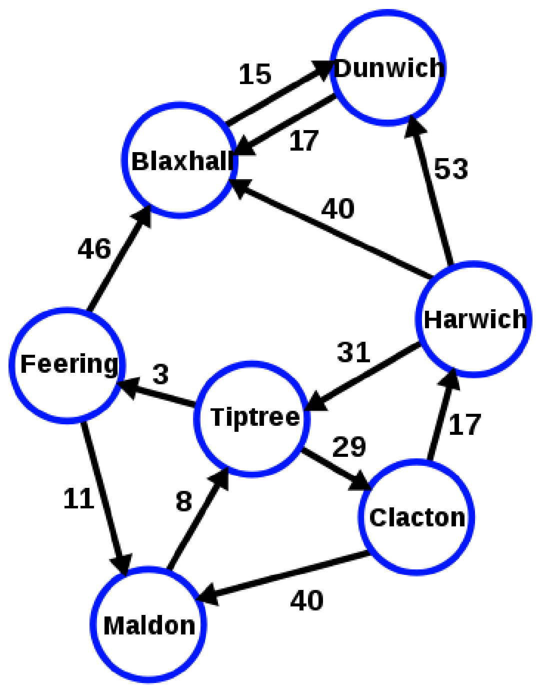
http://mathworld.wolfram.com/DirectedGraph.html

Graph theory: basics

Weighted graphs: $G_{\text{weighted}} \equiv \{V, E\}; \quad E \mapsto \mathrm{R}$

All edges carry a real (often positive) number, the weight. $f(e_{ij}) = w_{ij}$





Wikimedia

Graph theory: basics

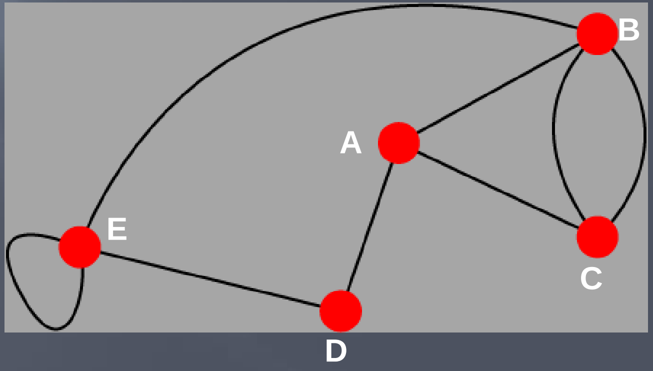
A path is a sequence of nodes in which each node is adjacent to the next one. $P_{0,n}$ of length n between nodes i_0 and i_n is an ordered collection of n+1 nodes and n links without repetition of links

$$P_{0n} = \{i_0, i_1, i_2, \ldots, i_n\}$$

- A path can intersect itself.

$$P_{0n} = \{e_{i_0 i_1}, e_{i_1 i_2}, e_{i_2 i_3}, \ldots, e_{i_{n-1} i_n}\}$$

- In a walk edges can be multiply visited. A walk on the graph on the right: **ABCBCADEEBA**



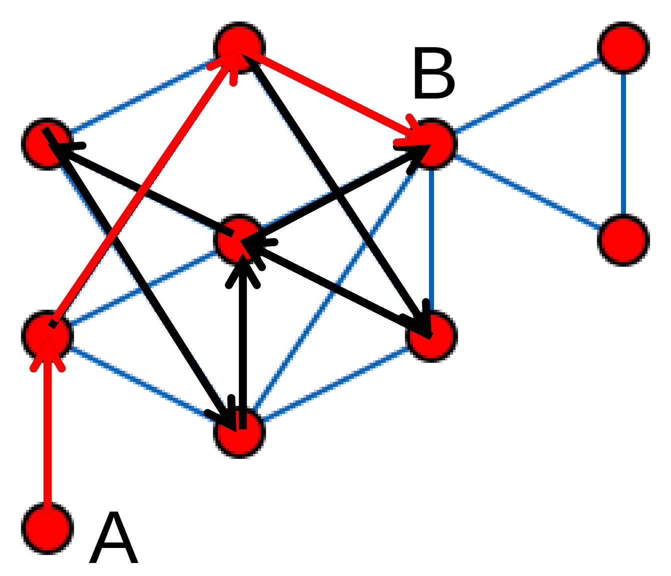
- A circle is a closed path (i_0=i_n)

- In a directed network, the path can follow only the direction of an arrow.

Barabási

Graph theory: basics

Distance: The length of the shortest path between two nodes. Length is measured in steps = # links.



Path length AB = 8

Distance d_{AB} = 3 (geodesic distance)

There can be more than one shortest paths.

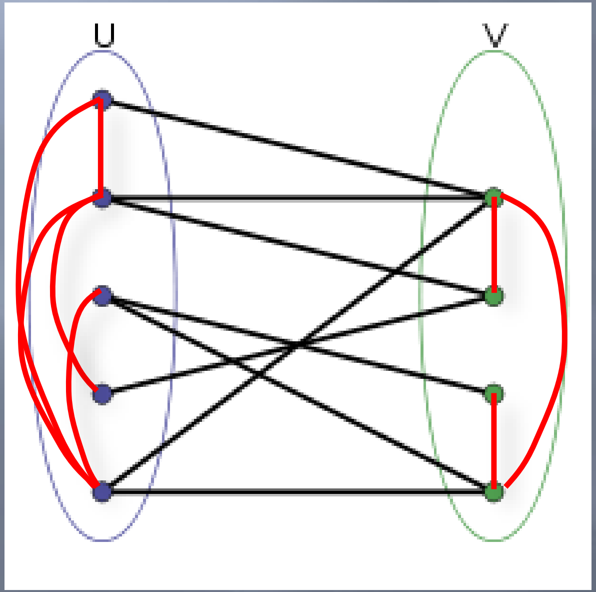
Wikimedia

Graph theory: basics

Bipartite graph:

$$G = \{U, V, E\}$$

$$e_{ij}\ \text{Î}\ E,\ \ i\ \text{Î}\ U,\ \ j\ \text{Î}\ V$$



Projections:

$$G_1 = \{U, E_1\}$$

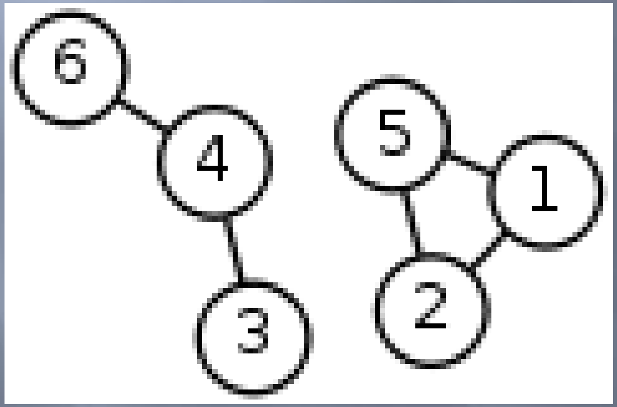
$$e_{ij}\ \text{Î}\ E_1 \text{ if } i, j\ \text{Î}\ U \text{ and } \ \$\ \{i, k, j\} \text{ path, } k\ \text{Î}\ V$$

$$G_2 = \{V, E_2\}$$

$$e_{ij}\ \text{Î}\ E_2 \text{ if } i, j\ \text{Î}\ V \text{ and } \ \$\ \{i, k, j\} \text{ path, } k\ \text{Î}\ U$$

Graph theory: basics

Graph 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.

Let *N* be the number of nodes n_s the number of components of size *s*.

Clearly $$N = \sum_{s=1}^{s_{\max}} s n_s$$

The concept of component is non-trivial for directed graphs, as the paths have to follow the arrows.

Wikimedia

Graph theory: basics

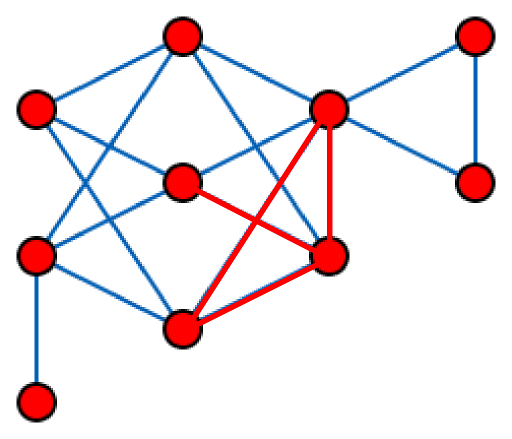
Subgraph of G: $G' = \{V', E'\}$ with $V' \subseteq V; E' \subset E$ such that

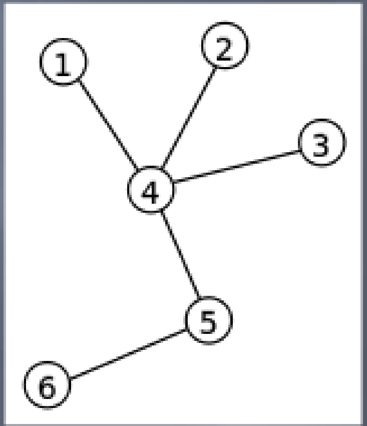
$$\forall e_{ij} \in E' \Rightarrow i, j \in V'$$

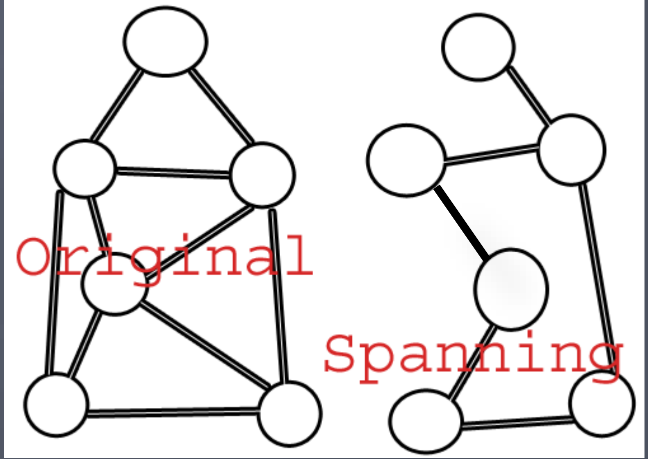
Spanning subgraph: $V' = V$

Tree: A graph with no circles (loops)

Spanning tree: A spanning subgraph with no loops



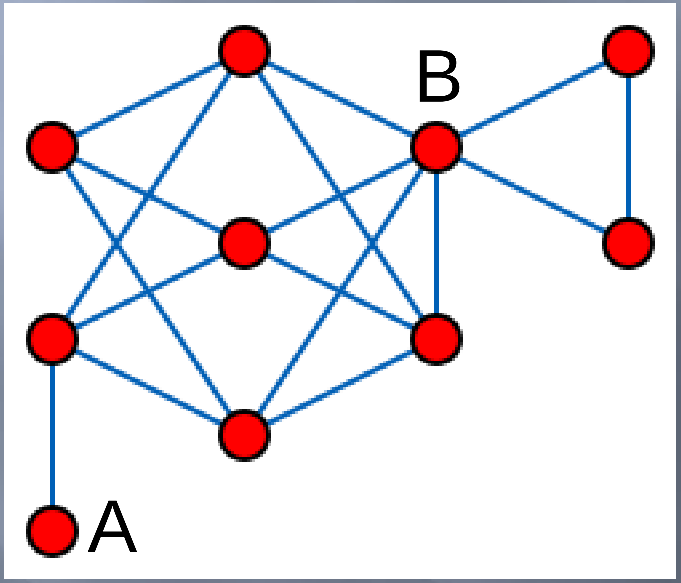


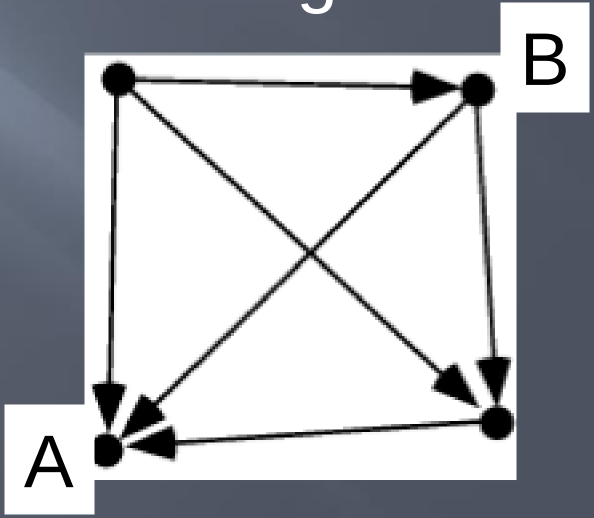


Graph theory: basics

Node degree k: The number of links from or to a node. For undirected it is the same.

For directed graphs: in and out degrees





$k_A = 1$
$k_B = 6$

$k_A^{in} = 3$ $k_A^{out} = 0$
$k_B^{in} = 1$ $k_B^{out} = 2$

Graph theory: basics

Distributions: In a large graph there are all kinds of nodes, the weights can be different etc.

Let us have a property x of the nodes, i.e., we have property x_i at node i. We can make a statistics over this property:

There are $n(x)$ nodes with property x
$n(x')$ nodes with property x' etc.

$n(x)$ is an important characterization of the system from the point of view of property x.

What is the average value of x?

Graph theory: basics

What is the average value of x?

$$\langle x\rangle = \frac{1}{N}\sum_{i=1}^{N} x_i = \frac{1}{N}\sum_{\forall x} xn(x) = \sum_{\forall x} xP(x)$$

where

$$P(x) = \frac{1}{N} n(x)$$

is the normalized (empirical) distribution of x

Average degree (L number of links):

$$\langle k\rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i = \frac{2L}{N}$$

$$\langle k^{in}\rangle \equiv \frac{1}{N}\sum_{i=1}^{N} k_i^{in} = \frac{1}{N}\sum_{i=1}^{N} k_i^{out} = \langle k^{out}\rangle = \frac{L}{N}$$

Undirected

Directed

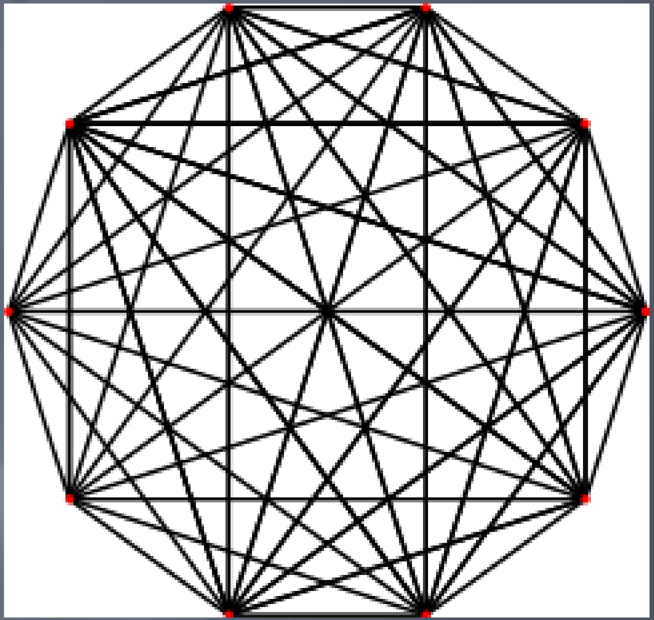
Graph theory: basics

Complete graph:

Simple graph with maximum number of links.

$$L = N(N - 1) / 2$$

$$k_i = N - 1 \quad \text{for} \quad " \; i$$



A complete graph is a regular graph: all nodes have the same degree and the graph is connected.

$$L \sim \mathcal{O}(N^{\lambda})$$

$\lambda = 1$ sparse graph (most cases)

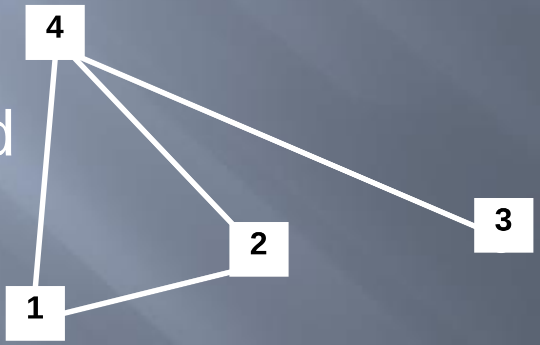
$\lambda = 2$ dense graph

Graph theory: basics

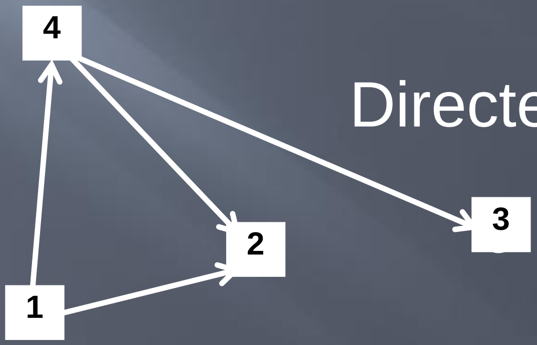
How to define a graph? Give a list of which nodes are connected.

The adjacency matrix A_{ij} is 1 if i is connected to j and 0 otherwise

Undirected



Directed



Symmetric

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

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

Non-symmetric

Graph theory: basics

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

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

$$k_j = \sum_{i=1}^{N} A_{ij}$$

Undirected

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

$$k_i^{out} = \sum_{i=1}^{N} A_{ij}$$

$$k_j^{in} = \sum_{i=1}^{N} A_{ij}$$

Directed

Graph theory: basics

Powers of the adjacency matrix $\boldsymbol{A}^n$:

$$(\boldsymbol{A}^2)_{ij} = \sum_k A_{ik} A_{kj}$$

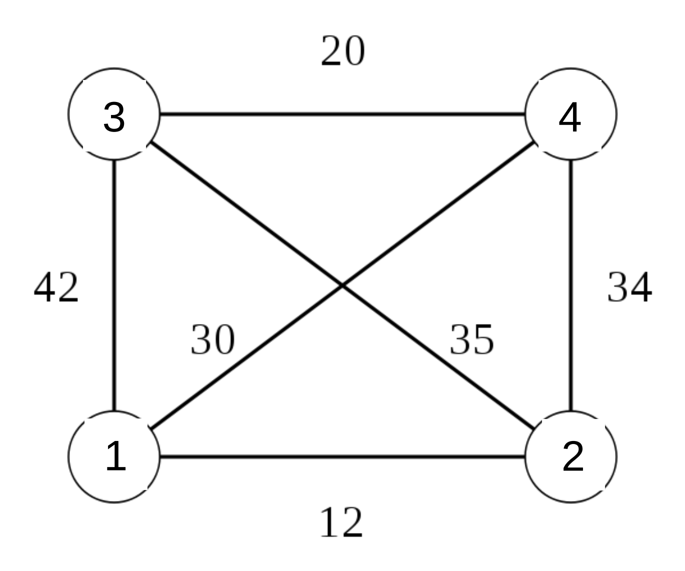
$$(\boldsymbol{A}^n)_{ij} = \sum_k (\boldsymbol{A}^{n-1})_{ik} A_{kj} \quad *$$

$(\boldsymbol{A}^n)_{ij}$ Gives the number of *n*-step walks (not paths!) 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.* All these cases are summed up in (*).

Graph theory: basics

Weighted graphs: adjacency matrix → weight matrix:



$$W_{ij} = \begin{pmatrix} 0 & 12 & 42 & 30 \\ 12 & 0 & 35 & 34 \\ 42 & 35 & 0 & 20 \\ 30 & 34 & 20 & 0 \end{pmatrix}$$

If undirected, still symmetric

Example of directed weight matrix

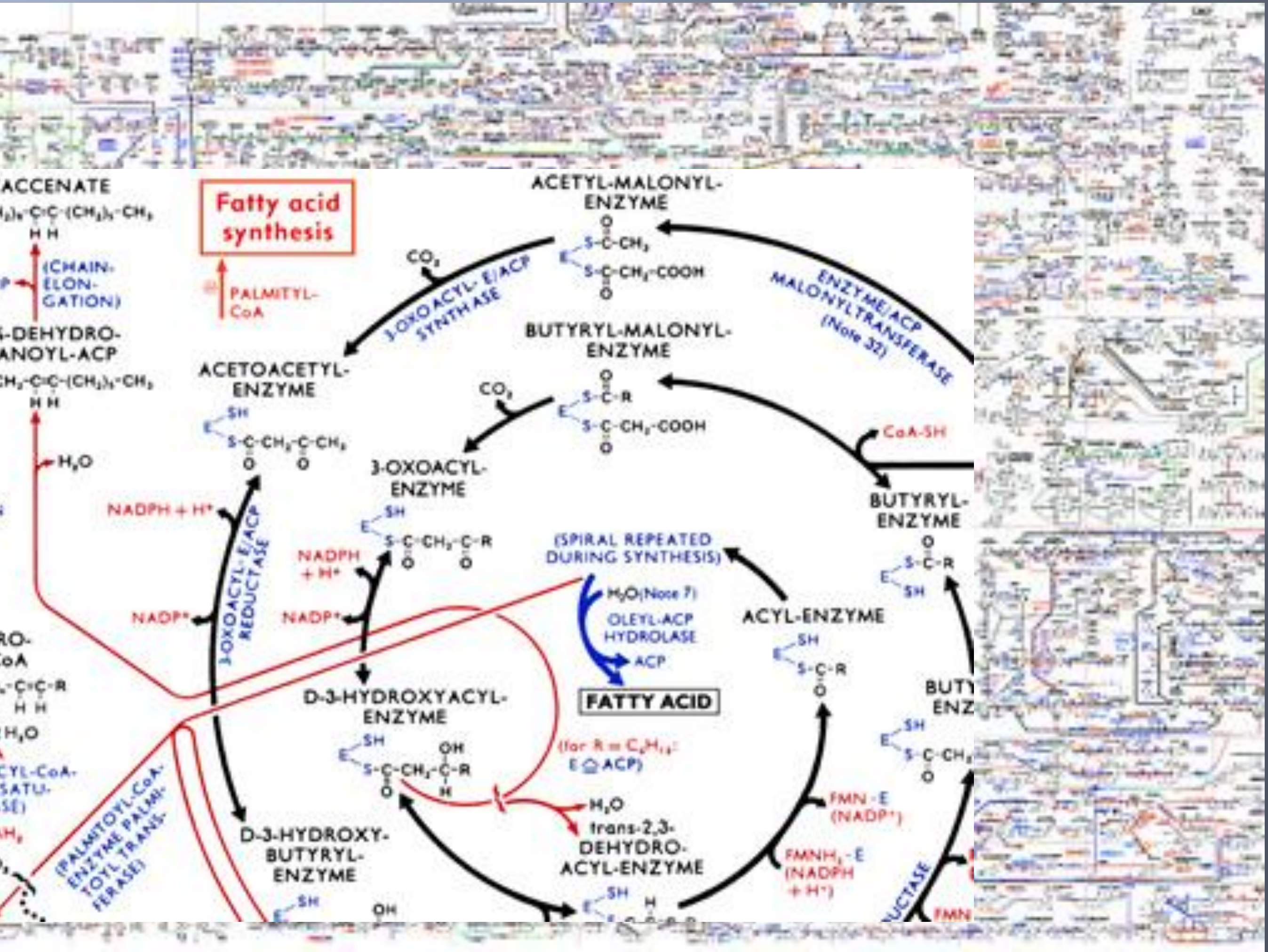
$$W_{ij} = \begin{pmatrix} 0 & 3.5 & 4.7 & 0 \\ 1.2 & 0 & 7.3 & 3.4 \\ 0 & 0 & 0 & 2.8 \\ 8.2 & 0 & 1.1 & 0 \end{pmatrix}$$

Graph theory: examples

Phenomenon	Nodes	Links
Cell metabolism	Molecules	Chem. reactions
Sci. collaboration	Scientists	Joint papers
www	Pages	URL links
Air traffic	Airports	Airline connections
Economy	Firms	Trading
Language	Words	Joint appearance

Graph theory: examples

Cell metabolism	Molecules	Chem. reactions



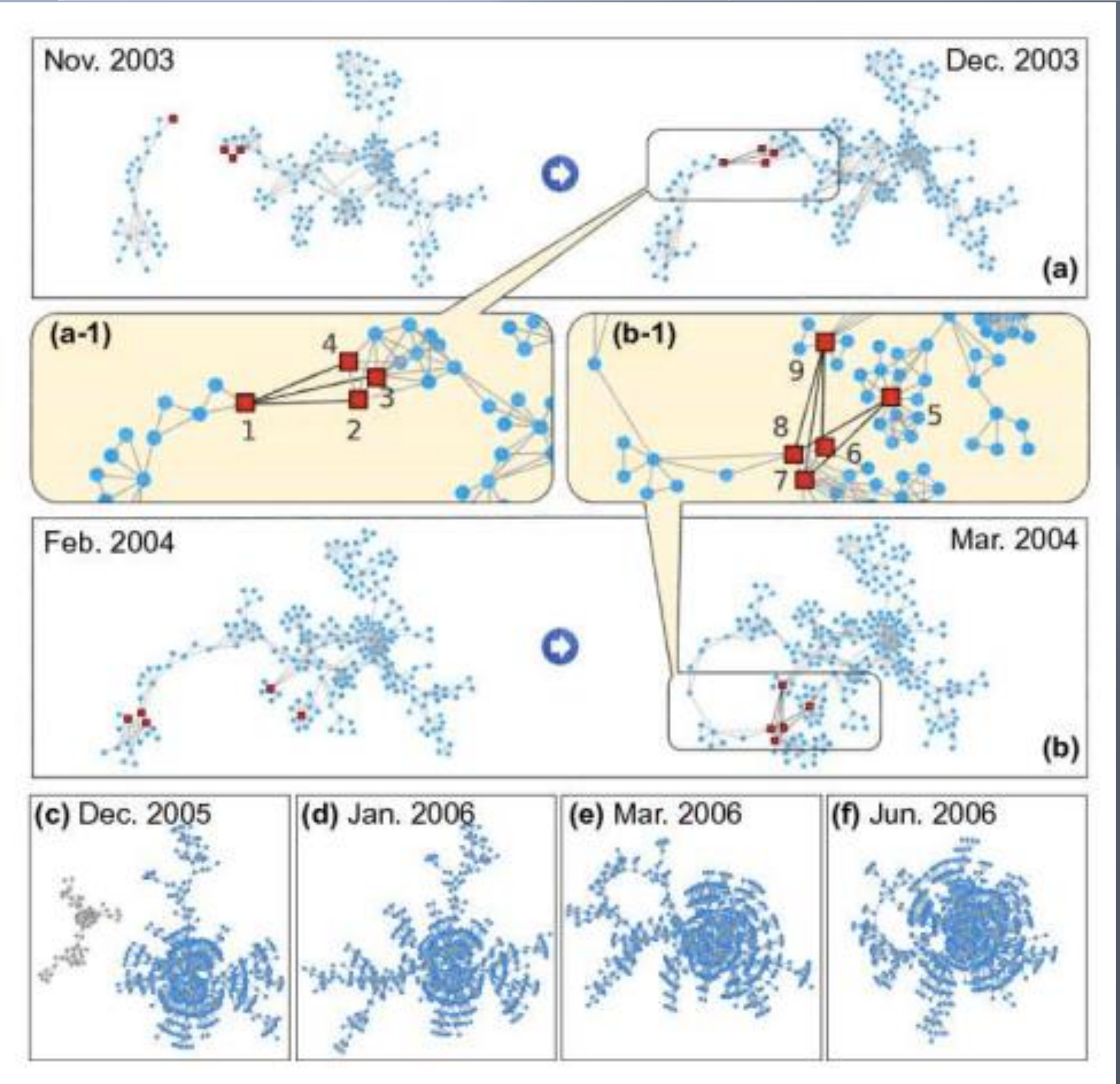
Undirected nw

(the arrows are for underlining the metabolic process.)

http://www.visualcomplexity.com/vc/project.cfm?id=224

Graph theory: examples

Sci. collaboration	Scientists	Joint papers



Bipartite graph:

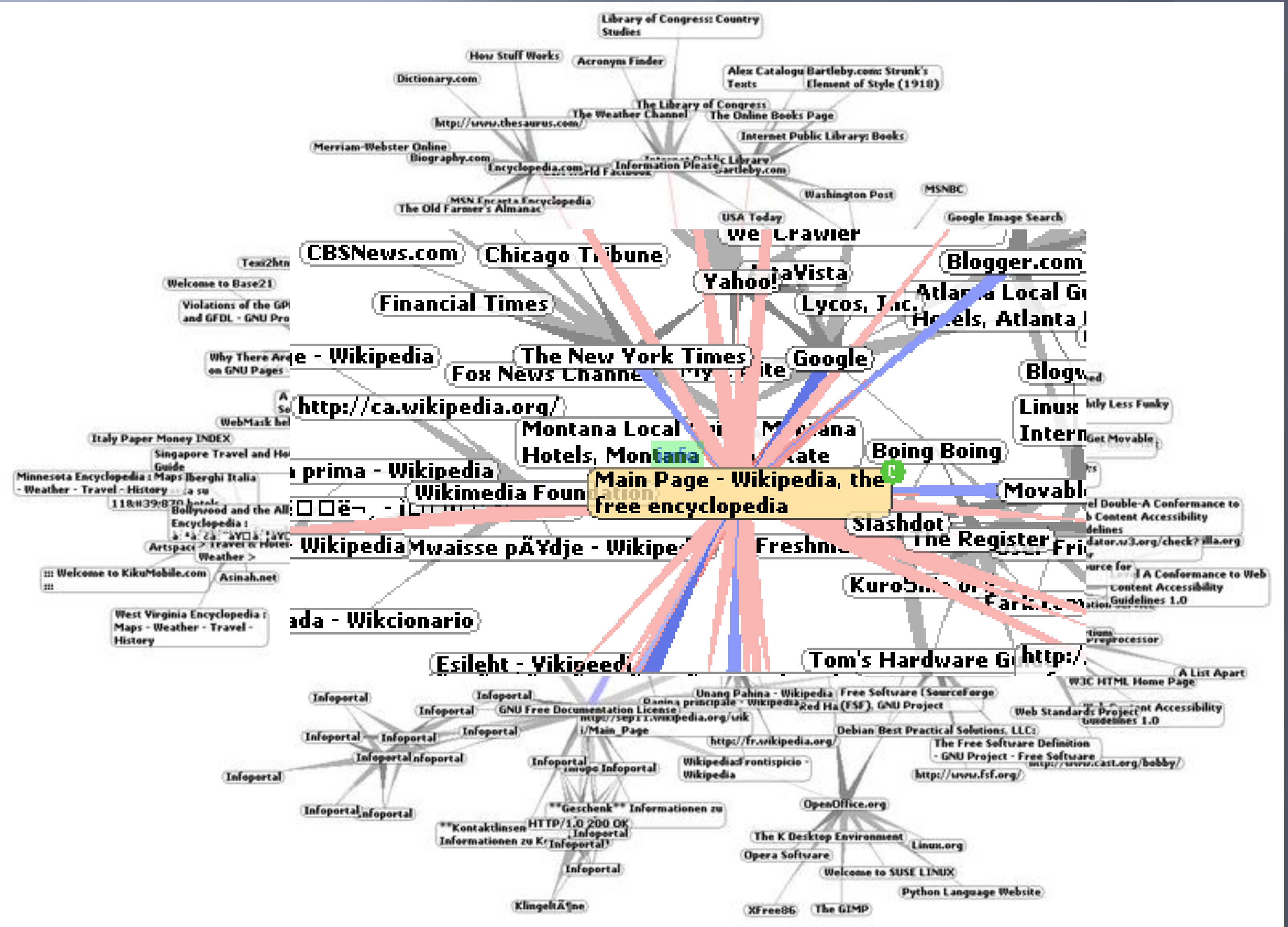
U: authors

V: papers

D. Lee et al. PHYS REV E , vol. 82, no. 2, 2010

Graph theory: examples

WWW	Pages	URL links



WWW arounnd Wikipedia main page

Directed network

Outgoing links

Wikimedia

Graph theory: examples

Air traffic	Airports	Airline connections	undirected
Economy	Firms	Trading	directed
Language	Words	Joint appearance	bipartite

Graph theory: important measures

1. Degree distribution $P(k)$

Given a network, the degrees of the nodes can take different values. If $n(k)$ is the number of nodes with degree k, the normalized distribution will be $P(k)=n(k)/N$. As for any normalized distribution

$$\sum_{k=0}^{k_{\max}} P(k) = 1$$

As discussed earlier:

$$\langle k \rangle = \sum_{k=0}^{k_{\max}} kP(k) = \frac{2L}{N}$$

An important characteristic for a distribution is the variance σ^2.

$$\sigma^2 = \left\langle (k - \langle k \rangle)^2 \right\rangle = \left\langle k^2 \right\rangle - \langle k \rangle^2 = \sum_{k=0}^{k_{\max}} k^2 P(k) - \left(\sum_{k=0}^{k_{\max}} kP(k) \right)^2$$

Always exists if $k_{\max} < \infty$

Graph theory: important measures

2. Average distance between nodes: This quantity is defined for a component (distance between components is infinite).

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

3. Diameter of a network:

$$\delta = \max_{(i,j)} d_{ij}$$

Usually, for N→oo

$$\langle d \rangle \sim \delta \sim N^{\lambda}$$

$\lambda = 0$, e.g., log: Small world

Graph theory: important measures

4. Clustering coefficient C_i at node i: What fraction of the neighbors of i are connected? Let the degree of i be k_i

Possible number of connections: $k_i\,(k_i - 1)/2$

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

where $n_\Delta(i)$ is the number of triangles at node i

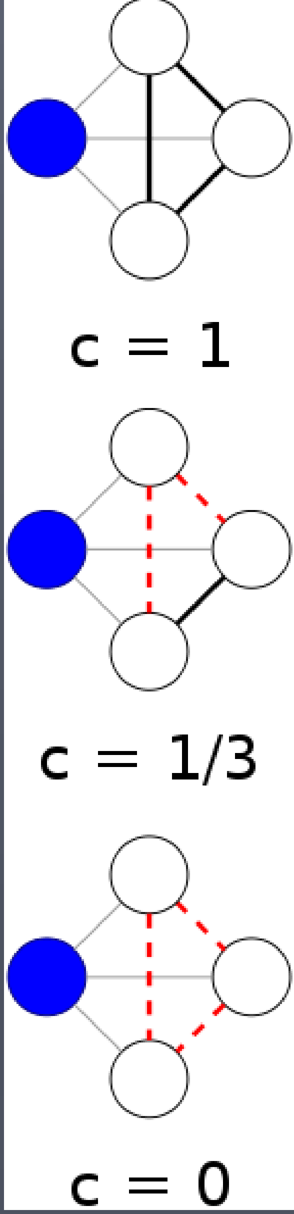
Average clustering coefficient

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

Global clustering coefficient:

$$C = \frac{\#\,\text{triangles} \times 3}{\#\,\text{connected triples}}$$

Note $C \neq \langle C \rangle$



Wikmedia

Graph theory: important measures

Conditional distribution: $P(x|\mathrm{cond.})$ is the normalized distribution of x, provided condition "cond." is fulfilled. Example: The clustering coefficients of nodes of degree k:

$$\langle C_k \rangle = \frac{1}{n_k} \sum_{i=1}^{n_k} C_i(k) = \sum_C C P(C|k)$$

5. Assortativity: The measure of the tendency that high degree nodes are neighbors of high degree

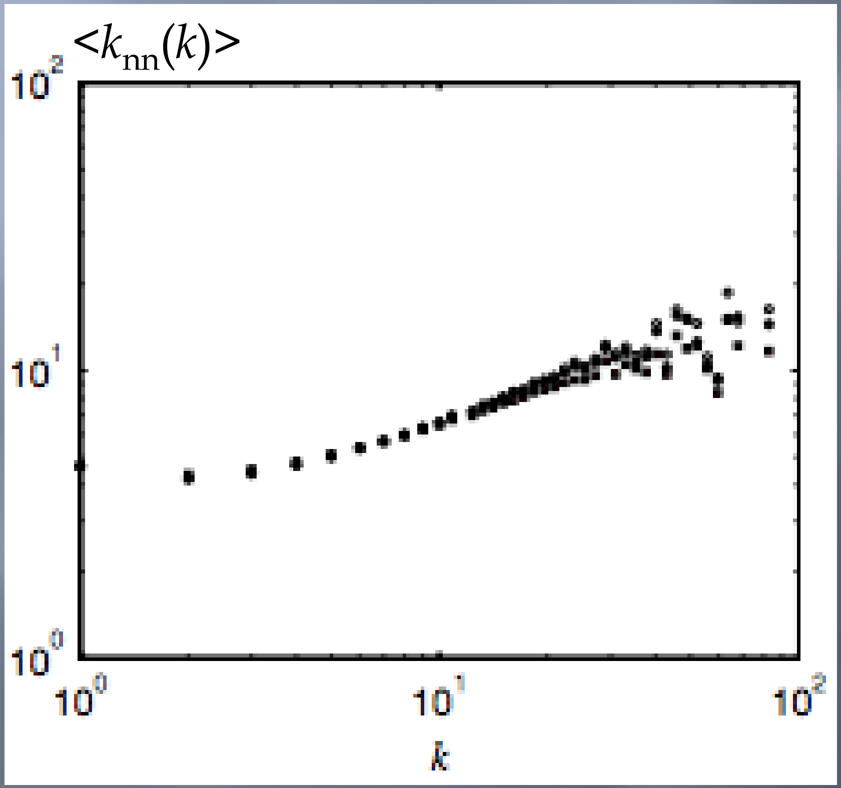
$P_{\mathrm{nn}}(k'|k)$ is the probability that a link from a node of degree k goes to a node of degree k'

$$\langle k_{nn}(k) \rangle = \sum_{k'} k' P_{\mathrm{nn}}(k'|k)$$

Is the expected degree of k-degree nodes.

Graph theory: important measures

If $\langle k_{nn}(k) \rangle$ is an increasing function of k, high degree nodes like to link to high degree nodes.



Mobile phone network

Assortative mixing

The opposite case is

disassortative mixing

Onnela et al. NJP, 9, 179 (2007)

Graph theory: important measures

How similar are two nodes *i* and *j*?
Jaccard coefficient:

$$J_{ij} = \frac{|N(i) \cap N(j)|}{|N(i) \cup N(j)|}$$

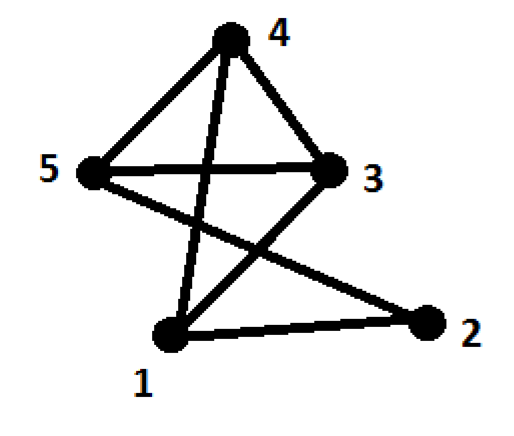
$$0 \leq J_{ij} \leq 1$$

where $N(i)$ is the set of neighbors of node *i.*

$J_{ij} = 0$ means entirely different

$J_{ij} = 1$ means equivalence

$$J_{15} = 1, J_{12} = 0, J_{13} = 1/5$$



Centrality measures

- If I need to recruit 10 people for my newly found organization, whom should I consider?
- If I am to pass on a message to three people in this network so that they in turn convey it to their friends and so on. Which three people should I select?
- If I am to rank all my friends based on how "central" they are in this network, how would I go about?
- If I were to nominate a leader for this team of 500, whom should I pick?
- How “important” is a node (link)?

Centralitiy measures

What makes a node (link) important?

1. Degree centrality High degree nodes are more important than low degree node *i*: k_i

Who has most connections?

2. Closeness centrality

$$C_s(i) = \left[\frac{1}{N-1}\sum_j d_{ij}\right]^{-1}$$

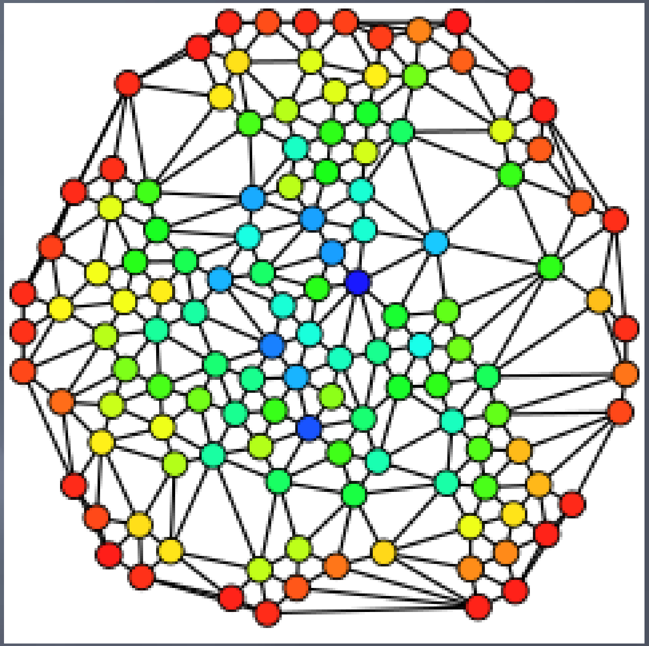
inverse of average distances between from *i*

Similarly: Harmonic centrality: inverse of harmonic mean (advantage: works for multi-componont graphs)

Who needs least effort to reach *everybody*?

Centrality measures

3. Betweenness centrality of a node (link): Calculate the fraction of shortest paths which go through that node (link). Sum it up over all pairs.



$$C_B(i) = \sum_{i \neq k \neq l} \frac{n_{kl}(i)}{n_{kl}}$$

Where are the bottlenecks?

Wikipedia

Centrality measures

4. Eigenvector centrality

Degree centrality is too simple. A node is important if it is connected to many important nodes. Give a score *x* to the nodes and calculate the new values:

$$x'_i = \sum A_{ij} x_j$$

If this is iterated (*x'*→*x*), a solution is found, which is related to the largest eigenvalue of A_{ij}. Let $\boldsymbol{v}_k$ the k-th eigenvector of ***A*** with eigenvalue λ_k, with max: λ_1

$$\mathbf{x}(t) = \boldsymbol{A}^t \boldsymbol{x}(0) = \boldsymbol{A}^t \sum_k c_k \boldsymbol{v}_k = \sum_k c_k \lambda_k^t \boldsymbol{v}_k =$$

$$\lambda_1^t \sum_k c_k \left(\frac{\lambda_k}{\lambda_1}\right)^t \boldsymbol{v}_k \to c_1 \lambda_1^t \boldsymbol{v}_1$$

meaning that

$$x_i = \left(\frac{1}{\lambda_1}\right) \sum A_{ij} x_j$$

Transmitted importance 1

Centrality measures

5. Katz-centrality

$$C_{\text{Katz}}(i) = \sum_{j} \sum_{k=1}^{\infty} \alpha^{k} \left(\boldsymbol{A}^{k}\right)_{ij}$$

$(\boldsymbol{A}^{k})_{ij}$ is the # walks btw *i* and *j*. The idea is that longer walks contribute less. To assure this, α < 1.

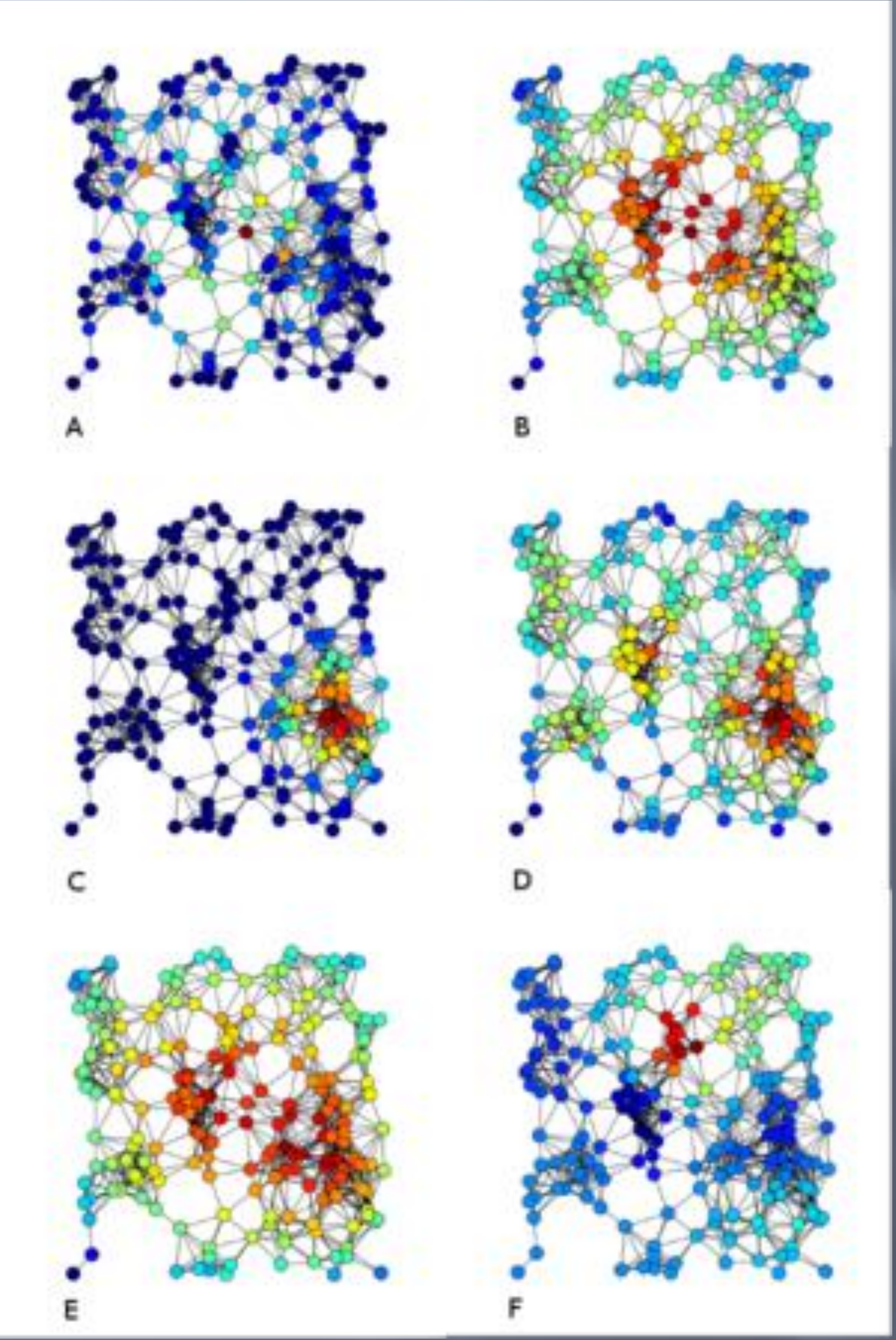
If $\alpha < 1/\lambda_1$, where λ_1 is the largest eigenvalue of ***A***, this formula is equivalent to:

$$C_{\text{Katz}}(i) = \sum_{j} [(\mathbf{1} - \boldsymbol{\alpha A})^{-1} - \mathbf{1}]_{ij}$$

Advantage: works also for directed networks

Transmitted importance 2

Centralities



A) Betweenness centrality
B) Closeness centrality
C) Eigenvector centrality
D) Degree centrality
E) Harmonic centrality
F) Katz centrality
of the same graph.

Wikipedia

How do real complex networks look like?

- **Small world**
- **Broad degree distribution**
- **High clustering**
- **Modular structure**

Universal features of many **very different** networks

Why?

How to model them? (Related questions)

Modeling networks

As technology advances we a) get access to b) generate large networks

We can easily generate regular networks (e.g., lattices) but in real networks there is usually a large amount of randomness.

Random network models will be the focus.

Home work

Take the Zachary karate club data (e.g., http://www-personal.umich.edu/~mejn/netdata/)

and calculate both the average clustering coefficient and the global clustering coefficient.