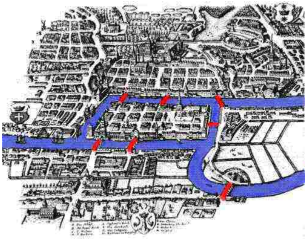
Networks and Graphs

Paul Van Dooren
Université catholique de Louvain
Louvain-la-Neuve, Belgium

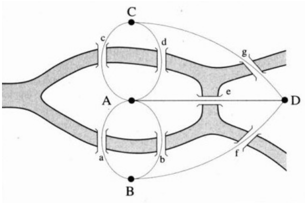
Large Graphs and Networks
UCL Université catholique de Louvain

Francqui lecture, May 21st 2010

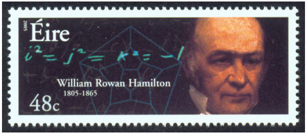
Graph theory started with Euler who was asked to find a nice walk across the seven Köningsberg bridges



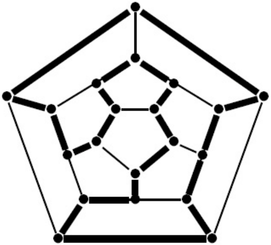
The (Eulerian) walk should cross over each of the seven bridges exactly once



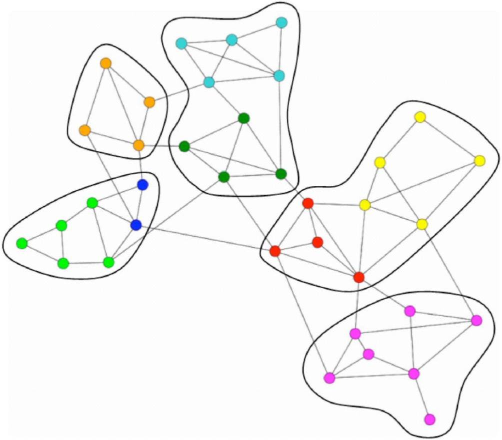
Another early bird was Sir William Rowan Hamilton (1805-1865)



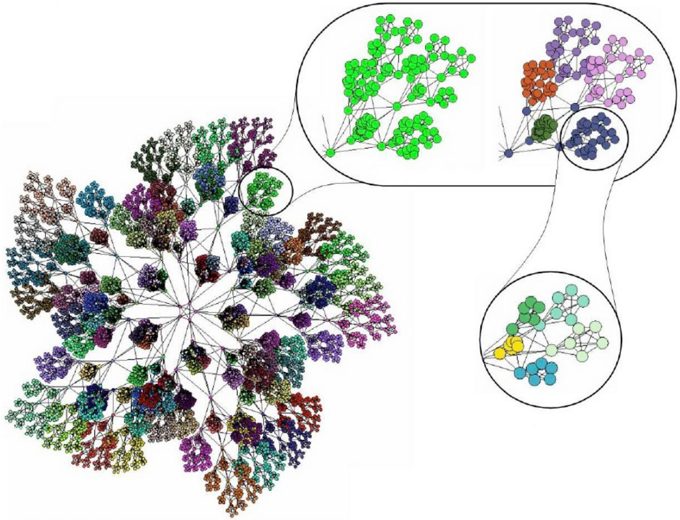
In 1859 he developed a toy based on finding a path visiting all cities in a graph exactly once and sold it to a toy maker in Dublin. It never was a big success.



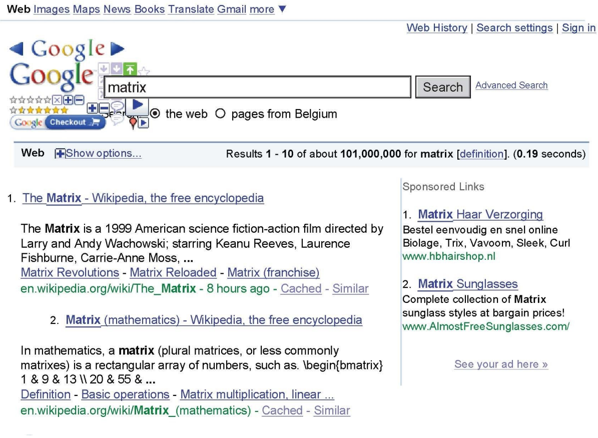
But now graph theory is used for finding communities in networks



where we want to detect hierarchies of substructures and their sizes can become quite big ...



It is also used for ranking (ordering) hyperlinks



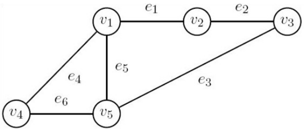
and for finding similarities between elements in a database

or by your GPS to find the shortest path home ...

What we will cover in this course

- Some basics about graphs
- Reminder of Perron-Frobenius
- A number of applications including :
    - ranking in a large graphs
    - similarity in large graphs
    - optimizing your PR
    - telephone network applications
    - clustering in large graphs

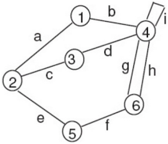
A graph $G = (V, E)$ is a pair of vertices (or nodes) V and a set of edges E, assumed finite i.e. $|V| = n$ and $|E| = m$.

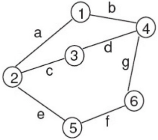


Here $V(G) = \{v_1, v_2, \ldots, v_5\}$ and $E(G) = \{e_1, e_2, \ldots, e_6\}$.

An edge $e_k = (v_i, v_j)$ is incident with the vertices v_i and v_j.

We focus on simple graphs (no self-loops or multiple edges) :





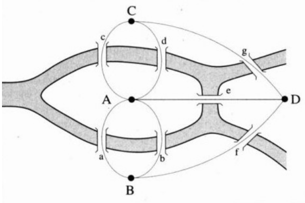
Some properties

The degree $d(v)$ of a vertex V is its number of incident edges

A self-loop counts for 2 in the degree function.

The sum of the degrees of a graph $G = (V, E)$ equals $2|E|$

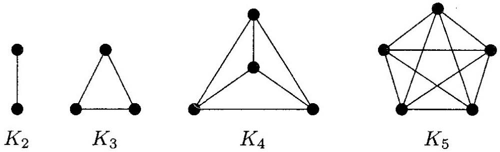
Corollary The number of vertices of odd degree is even (Euler !)



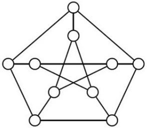
Euler walks exist iff there are at most two vertices of odd degree

Special graphs

A complete graph K_n is a simple graph with all $B(n, 2) := \frac{n(n-1)}{2}$ possible edges, like the matrices below for $n = 2, 3, 4, 5$.



A k-regular graph is a simple graph with vertices of equal degree k



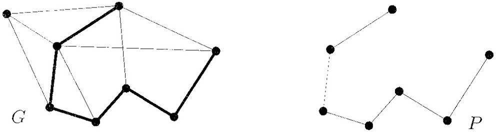
Walking in a graph

A walk of length k from node v_0 to node v_k is a non-empty graph $P = (V, E)$ of the form

$$V = \{v_0, v_1, \dots, v_k\} \quad E = \{(v_0, v_1), \dots, (v_{k-1}, v_k)\}$$

where edge j connects nodes $j - 1$ and j (i.e. $|V| = |E| + 1$).

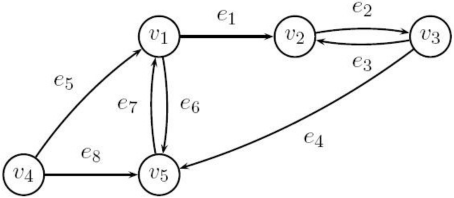
A path is a walk with all different nodes (and hence edges).



A cycle is a walk with different nodes except for $v_0 = v_k$.

Directed graphs

In a directed graph or digraph, each edge has a direction.



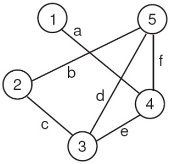
For $e = (v_s, v_t)$, v_s is the source node and v_t is the terminal node.

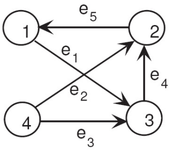
Each node v has an in-degree $d_{in}(v)$ and an out-degree $d_{out}(v)$.

Representing graphs

A graph $G = (V, E)$ is often represented by its adjacency matrix.

It is an $n \times n$ matrix A with $A(i,j) = 1$ iff $(i,j) \in E$. For the graphs



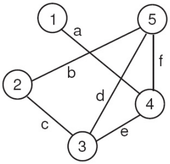


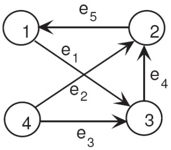
the adjacency matrices are

$$A_1 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 1 & 1 & 0 \end{bmatrix} \quad A_2 = \begin{bmatrix} 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 1 & 0 \end{bmatrix}$$

A graph can also be represented by its $n \times m$ incidence matrix T.

For an undirected graph $T(i,k) = T(j,k) = 1$ iff $e_k = (v_i, v_j)$.
For a directed graph $T(i,k) = -1$; $T(j,k) = 1$ iff $e_k = (v_i, v_j)$.
For the graphs

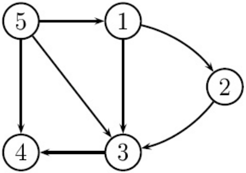




the incidence matrices are

$$T_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 0 & 1 \end{bmatrix} \quad T_2 = \begin{bmatrix} -1 & 0 & 0 & 0 & 1 \\ 0 & 1 & 0 & 1 & -1 \\ 1 & 0 & 1 & -1 & 0 \\ 0 & -1 & -1 & 0 & 0 \end{bmatrix}$$

One can also use a sparse matrix representation of A and T.
This is in fact nothing but a list of edges, organized e.g. by nodes.



$$V(1) = \{2, 3\}$$
$$V(2) = \{3\}$$
$$V(3) = \{4\}$$
$$V(4) = \emptyset$$
$$V(5) = \{1, 3, 4\}$$

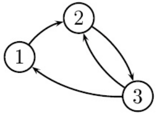
Notice that the size of the representation of a graph is thus linear in the number of edges in the graph (i.e. in $m = |E|$).

Powers of A

Proposition $(A^k)_{ij}$ is the number of walks of length k from i to j

Proof Trivial for k=1; by induction for larger k.
The element (i, j) of $A^{k+1} = A^k \cdot A$ is the sum of the walks of length k to nodes that are linked to node j via the adjacency matrix A.

One verifies this in the following little example



$$A = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{bmatrix}, \quad A^2 = \begin{bmatrix} 0 & 0 & 1 \\ 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix}$$

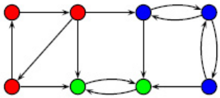
Corollary In a simple undirected graph one has the identities
$tr(A) = 0, tr(A^2)/2 = |E|$ and $tr(A^3)/6 = |\text{triangles in } G|$.

Exercise What is the complexity of counting this ?

Connected components

In a directed graph $G = (V, E)$, u and v are strongly connected if there exists a walk from u to v and from v to u.

This is an equivalence relation and hence leads to equivalence classes, which are called the connected components of the graph G.



The graph reduced to its connected components is acyclic (why ?)

This shows up in many applications, e.g. in the dictionary graph.

The connected components are the groups of words that use each other in their definition (see later).

Perron-Frobenius

Proposition Let A be a non-negative matrix. Then the spectral radius $\rho(A) := \max_i |\lambda_i|$ of A is also an eigenvalue of A.

Unicity
If A is irreducible then its multiplicity is 1; the corresponding eigenvector x is “unique” and strictly positive (PageRank !)

Convergence
If, moreover, the matrix is primitive, i.e. GCD(cycle-lengths)=1, then the second eigenvalue is strictly less than $\rho(A)$

A non-negative matrix A is irreducible if there does not exist a permutation P such that $P^T AP$ is block triangular.

The adjacency matrix A of a graph is irreducible iff the graph is strongly connected

Some references

J.A. Bondy and U.S.R. Murty, *Graph Theory with Applications*, North Holland, 1976.

Reinhard Diestel, *Graph Theory*, Graduate Texts in Mathematics, Vol. 173, Springer Verlag, Berlin, 1991.

Douglas West, *Introduction to Graph Theory*, Prentice Hall, 2000.

V. Blondel, A. Gajardo, M. Heymans, P. Sennelart, and P. Van Dooren. A measure of similarity between graph vertices : applications to synonym extraction and web searching. SIAM Review, 46(4):647-666, 2004.

C. de Kerchove, L. Ninove, P. Van Dooren, Influence of the outlinks of a page on its PageRank. Linear Algebra and its Applications, 429(5-6):1254-1276, 2008.

P. Van Dooren, C. Fraikin, Similarity matrices for colored graphs. Bull. Belg. Math. Soc. Simon Stevin, 16:705-722, 2009.