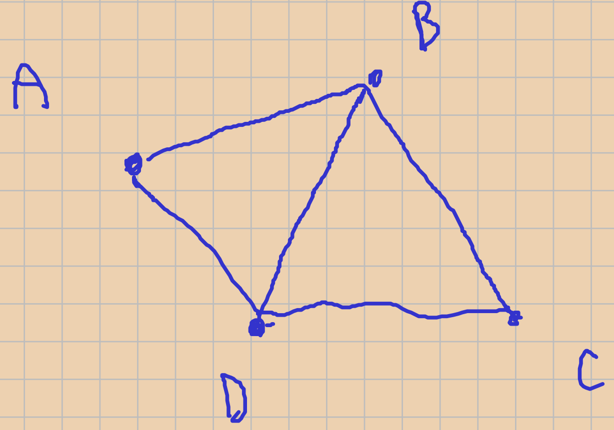
LT02

We talk about metrics

We talk about 3 main metrics:

- Degree
- Betweenes
- average path

A graph is a collection of vertices (nodes) and edges (links), it's a mathematical object $G(V,E)$



$$V = \{A, B, C, D\}$$

$$E = \{\{A,B\}, \{B,C\}, \ldots\}$$

4 Vertices and 5 Edges

$\{A,B\} = \{B,A\}$ in this example

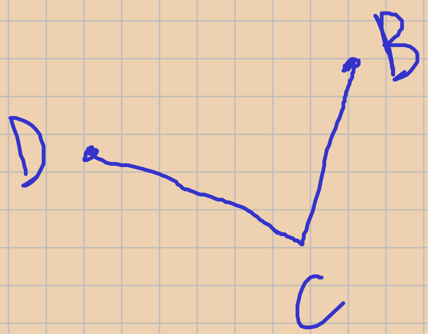
Order: the number of vertices

Size: the number of edges

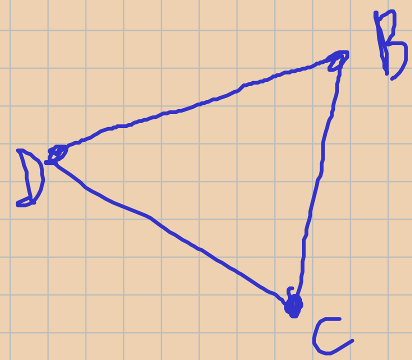
A subgraph of $G(V_g, E_g)$ is $H=(V_h, E_h)$

An induced subgraph is a subgraph when, once the nodes selected are chosen, we need to import ALL the edges related to those nodes.

ex



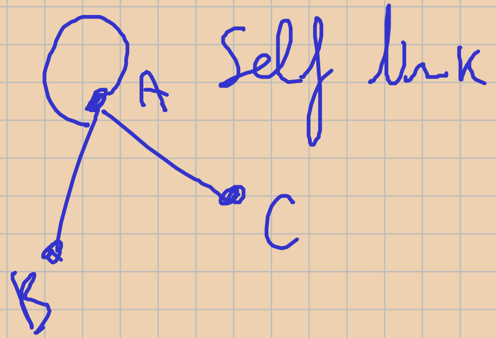
subgraph

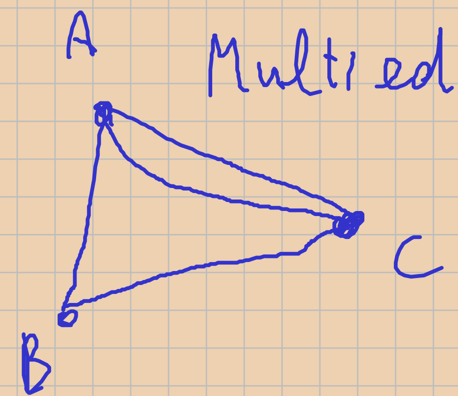


induced subgraph

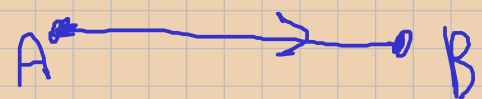
A graph with self-links (so a link comes out of A and comes back in A) or multi edges (where two nodes have more than 1 link between them), are called multi-graph or multiplex

In this course we will almost never treat multi-graphs.





A graph where the relationship between vertex A and B presents a directionality is called a directed graph



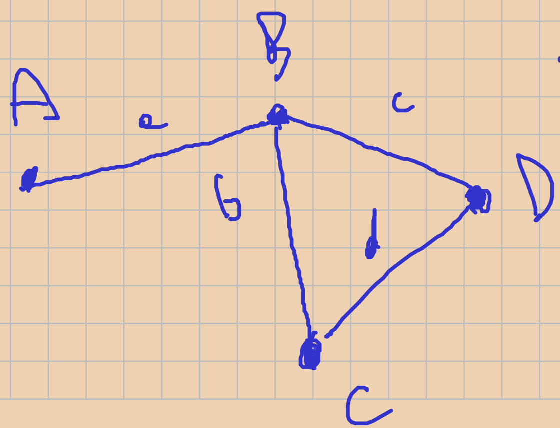
So : $\{A, B\} \neq \{B, A\}$

walk : a walk is a collection of nodes and links, the "lenght" is the number of links the walk has.

If we wanna go from one node to another, in a complex network, we probably have multiple way to do it, so we have multiple possible walks.

$$\{\underbrace{\nu_0}_{\text{node}}, \underbrace{l_1}_{\text{link}}, \nu_1, l_2, \dots, \nu_{n-1}, l_n, \nu_n\}$$

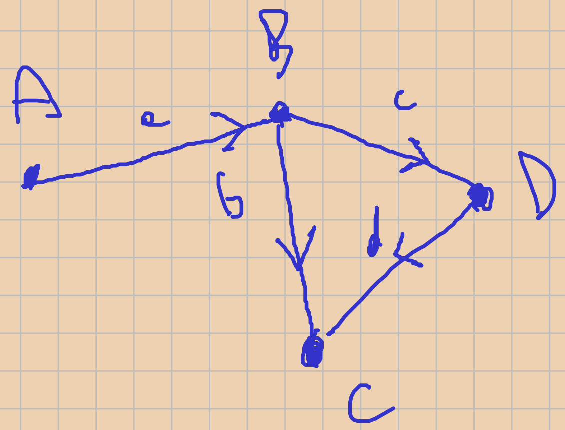
l_x



W : AaBcD

length = 2

If we have a directed graph, a Walk and the exact reverse Walk are not possible, because two nodes might be connected in a direction, but not in the other.



Here the same W as before can be done, but the reverse is not possible

Trial: a walk without repeated edges(links)

Path: a trial without repeated vertex(nodes)

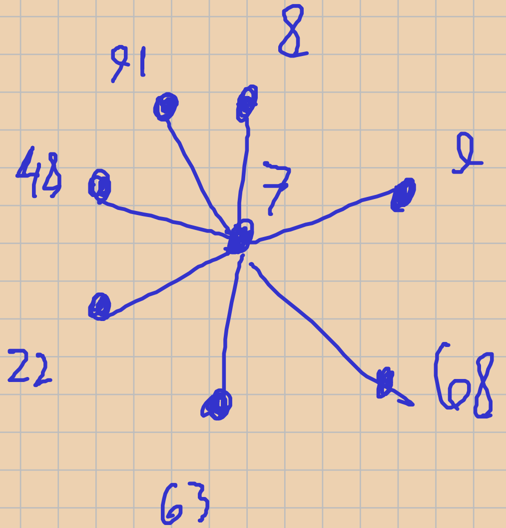
So a walk with no repeated nodes and links is called path.

A walk with miminum lenght 3 where the starting and ending nodes are the same is called a cycle.

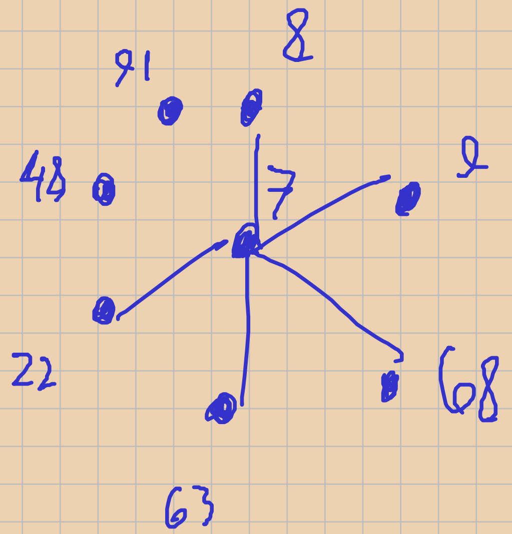
A vertex v is "reachable" from v if there exist at least a walk connecting u and v

A graph G is said to be "connected" if every vertex is reachable from every other one.

A component of a graph is a subgraph of it, that is "maximally connected", so every node in it is reachable.



Component



Not a Component

We also have a lot of other definitions, we can come back to these slides when we need it later in the course.

Degree: the degree of a node is the number of incident links on it.

Degree sequence: obtained by arranging the degree of nodes (in a complex networks), in increasing order

ex: $\{1, 1, 1, 1, 2, 2, 2, 2, 2, 2, 3, 3, 4\}$

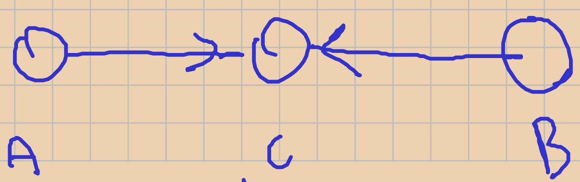
This means we have 4 nodes with degree 1, 5 with degree 2, 2 with degree 3, and just one node with degree 4.

We often write another sequence just under this, with the name of the nodes, in this order.

ex: $\{\underbrace{A, B, F, K}_{\text{1 degree}}, \underbrace{Z, C, R, S, V}_{\text{2 degree}}, \dots, \dots\}$

In a "digraph" (a directional graph), we have "in-degree" and "out-degree"

ex:

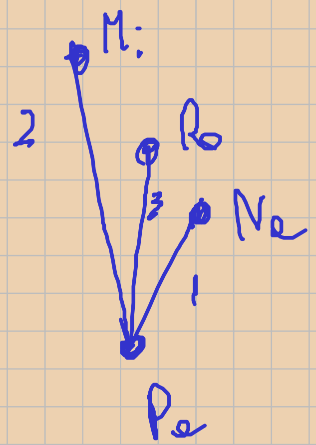


Outdegree = 0
indegree = 2

The sum of in-degree and out-degree gives always the degree.

We can have in-degree sequences and out-degree sequences.

Strenght: is just like degree, but we consider the weight of the links (one link might be more significative than others) the strenght of a node is the sum of the weight of all the links connected to the node.



Pe : Degree = 3
Strenght = 6

Degree is a centrality metric.

We introduce another centrality metric, betweennes: the ratio of 2 numbers: to compute the betweennes of one specific node

I take another 2, generic, nodes: we examine the number of paths (no repetition of nodes and links), and we only keep the shortests. We make a ratio between the ones that pass trought the first node (the one we want to calculate betweennes) and the total number of shortests paths. If we do that for every possible couple of nodes, hat is the betweennes.
A degree with high betweennes is a vulnerability of the system.

diameter: is a metric, the diameter of a network is the lenght of the biggest of the shortests paths. We compute all the possible shortests paths between every couple of nodes connected, from all these, the lenght of the longest one is the diameter.

Local clustering coefficient and global clustering coefficient

$$\underbrace{C_i}_{\text{Loc. clust. coeff. of the node } i} = \frac{2\,|\text{number of links observed between pairs of nodes(nodes connected to } i)|}{\underbrace{k_i}_{\text{degree of the node } i}(k_i - 1)}$$

$$\frac{k(k-1)}{2}$$

We will se this quantity a lot

Is a measure of how well connected are the nodes close to the one I'm interested in. It has values spanning from 0 to 1.

Assortativity: it's a preference for nodes with high degree to be connected with nodes with high degree, the assortativity of a network is the measure of this.

A network can be assortative or disassortative, if it has high or low assortativity