Lecture 7

Networks Everywhere
My distance from the Pope is: 5

A friend of the mother of a friend of mine Luca works in the Vatican city and he knows well the head of the Swiss grads who meets the Pope regularly.

Your distance from The President of United States of America?
A network (or graph) is a set of vertices connected via edges.

Vertices are sometime called nodes (or “agents”)

Edges are sometime called links (or “interactions”)

The number of edges ($E$) is called graph measure.

The notation $G(V,E)$ represents a graph with $V$ vertices and $E$ edges.

Examples of networks?
What do we want to know?

- Minimal travel length between two nodes.
- Average travel length in the network.
- Maximal travel length in the network.
- Distribution of individual connectivity.
- Vulnerability and resilience.
- Information flow.

How can we compute these properties, and can we infer some of them without the complete knowledge of the network structure?
Networks are not all the same

For instance, there is one fundamental difference between the internet and the predator-prey networks.

Edges can be directed, in this case the node ‘i’ can be connected with vertex ‘j’ but not vice versa

![Directed Graph Example](image)

Graph containing directed edges are called directed graphs, vice versa graphs without directed edges are undirected graphs.

![Undirected Graph Example](image)
Edges might carry “weights” (such as the probability or success rate of being crossed), in such case they are called **weighted graphs**.

*We encountered already weighted and directed graphs in the statistical mechanics approach where the probability to jump from one state to the other was* $W_{ij}$.
A **loop** is an edge that connects the vertex to itself

![loop diagram](image)

**Multiple edges** connect the same two vertices

![multiple edges diagram](image)

**Simple graphs** contain no loops or multiple edges

*Note that sometime one can use multiple edges to introduce weights in un-weighted graphs. Note also that often the term ‘loop’ is used to describe generic paths that come back to the starting point (however the correct term should be cycle).*
A **walk** is an alternating sequence of vertices and edges starting and ending with a vertex.

![](image1)

A **cycle** is a closed walk where the starting and ending vertex are the same.

![](image2)

A (simple) **path** is a walk where no vertices (and thus no edges) are repeated.

The **length** of a walk (or of a path) is the number of edges that it uses.

A **simple cycle** is a cycle where no vertices are repeated - except for the starting/ending one.

*(Often cycles are called loops.)*
A **Hamiltonian path** is a path that passes once through all the vertices (not necessarily all the edges).

A *Hamiltonian cycle* is a Hamiltonian path which starts and ends in the same vertex.

An **Euclidean path** is a path that passes once through all the edges (not necessarily all the vertices).

An *Euclidean cycle* is a Euclidean path which starts and ends in the same vertex.
A Graph is **connected** if a path exists for any couple of vertices in the graph.

A **dominating set** for a graph is a set of vertices whose neighbors, along with themselves, constitute all the vertices in the graph.
Some cycles can be reduced into smaller cycles.

A cycle is *reducible* if exists a path connecting any two vertices in the cycle which is shorter that the path through the edges belonging to the cycle.

*There exist “Shortcuts” within the set of nodes belonging to the cycle.*
The **degree** (or connectivity) of a node $k_i$ is the number of edges incoming and/or outgoing from the node.

The number of edges incoming into a node is the **in-degree**.

The number of edges outgoing from the node is the **out-degree**.

The **neighbors** of a given vertex ‘$i$’ are all the vertices ‘$j$’ connected with ‘$i$’ through et least one edge.

The number of neighbors of a given vertex is its degree.
The **degree distribution** is the probability distribution (relative frequencies) of the degrees in a given graph.

\[
p(k) = \frac{\text{# of vertices with degree } k}{\text{total # of vertices } (= V)}
\]

A graph where all vertices have the same degree is called **regular** (*regular graphs are not necessarily lattices or ‘ordered’*).
A connected graph with no cycles is a **tree**

![Diagram of a tree]

In a tree, the relation $V - E = 1$ always holds

A disjoint union of trees is a **forest**

![Forest diagram]
Jumping on trees
Let us investigate the sequence of the number of nodes that we encounter starting from a given node and moving outwards.

Coordination sequence,
or “shell map”.

\[ V(t + 1) = \sum_{i \in \{t \text{ shell}\}} (k_i - 1) \]

\[ V(t + 1) = \sum_k (k - 1) \frac{k \ p(k)}{\langle k \rangle} V(t) \]

with \( \lambda = \log \left( \frac{\langle k^2 \rangle}{\langle k \rangle} - 1 \right) \)

Exponential growth \( \Rightarrow \) hyperbolic space!

The **distance** \((d_{i,j})\) between two nodes in a graph (or topological distance) is the length of the shortest path between the two nodes.

The **diameter** of a finite graph is the maximum distance in the graph.

\[
\text{diameter} = \max_{i,j} (d_{i,j})
\]

If \(V(t) \propto \exp(\lambda t)\) then the diameter \((t_{\text{max}})\) is \(\approx \log(V)\)
The **complete graph** \((K_V)\) is a graph where every vertex is connected (one-edge path) with all the other vertices.

The number of edges in a complete graph is
\[
E = \frac{V(V-1)}{2}
\]

The degree of each vertex is \(V-1\).

The diameter of a complete graph is 1, independent on \(V\).

The number of triangles (three-cycles) is
\[
= \frac{V(V-1)(V-2)}{6}
\]

A **clique** is a set of vertices which are all connected to each other.

A *complete bipartite clique* \(K_{i,j}\) is a graph where every one of \(i\) nodes has an edge directed to each of the \(j\) nodes.
The **clustering coefficient** $C_i$ is the number of edges between the vertices which are neighbors of vertex ‘$i$’ divided by the total number of edges that could exist between these vertices.

\[
C_i = \frac{\text{# of edges between the neighbours of } i}{k_i(k_i - 1)/2}
\]

In the complete graph the clustering coefficient is 1.

\[
C_i = \frac{3}{21} = 0.14
\]
Average clustering coefficient (often called clustering coefficient)

\[ C = \frac{1}{V} \sum_i C_i \]

A different definition (often used):

\[ \tilde{C} = 6 \frac{\# \text{ triangles in the graph}}{\text{number of paths of length 2}} \]

Trees have clustering coefficient(s) = 0
**Betweenness**

The betweenness (centrality) of a vertex $i$ is the number of geodesic paths between other vertices that run through $i$.

$$B_i = \sum_{k > j \neq i} \frac{\# \text{ of geodesic paths between } k \text{ and } j \text{ passing through } i}{\# \text{ of geodesic paths between } k \text{ and } j}$$

*The betweenness can be seen as a measure of the network resilience to targeted attacks on vertices.*
Why the world is so small?
Calculate the diameter of the network below for the two cases:
1) ignoring the “red” edges; 2) including the red edges