1. Networks

A network (also called a graph) is a set of **nodes** connected by **edges**.

- **Nodes** → components that make up the graph (also called as "points" or "vertices").
- **Edges** → links that connect nodes together (also called as "links" or "connections").

Types of Networks

Networks can be divided into types based on:

1. Link Direction → **Undirected** or **Directed**
2. Edge Type → **Unweighted** or **Weighted**
3. Mode → **Bipartite** or not
4. Topological Evolution → **Static** or **Evolving**

Applications of Networks

- **Technological** → WWW, Internet, Electricity Grid, Routers, Neural networks
- **Social** → Friendships, Contracts, Interactions, Covert networks
- **Biological** → Epidemic/Disease, Species, Gene Regulation, Protein Interactions
- **Transportation** → Air, Sea, Railway, Highway networks

Representation of Networks

1. **Adjacency Matrix**: Create a matrix with all the nodes as the columns and rows. Write a 1 if the nodes are connected, otherwise 0.
2. **Edge List**: Make a list of all edges by writing which nodes they connect.
3. **Adjacency List**: Make a list of which node each node is connected to.

1.1. History of Networks

The foundation of Network Science is in **Graph Theory**, which was "started" by **Leonard Euler** in 1741 when he solved the problem of the **seven bridges of Königsberg**.

The Seven Bridges of Königsberg problem

There are 4 pieces of land connected by 7 bridges. Find a way to go through every bridge once without overlapping.

Euler proved that it is mathematically impossible to do this

By early 20th century, **social scientists** saw the practical uses of networks. **Sociograms** were made by a sociologist, *Jacob Moreno* which were an analysis of "social relations".

Six Degrees of Separation

To connect ANY two random persons in this world you only need a chain of 6 people.

This was an experiment done by the social psychologist, **Stanley Milgram**. This finding contributed a lot to the concept of the "small world".

1.2. Metrics/Measures

A metric is just a number used to measure a quantity. We can have 2 types of metrics/measures in a network:

1. **Local Metrics** → Metrics that apply on a per-node level, to each node individually.
2. **Global Metrics** → Metrics that apply to the entire network as a whole

There are two main local metrics in networks:

> **Degree** → The total number of edges directly connecting a node with other nodes in the network is that node's degree. It can be calculated via an **Adjacency Matrix**.

> **Strength** → The total value of the sum of all edges directly connecting a node with other nodes in the network is that node's degree. It is also called a **weighted** degree since it only applies in weighted networks.

In directed networks, we can also have **Indegree** and **Outdegree**:

- **Indegree** → No of edges that come from A to B would be A's indegree.
- **Outdegree** → No of edges that go from A to B would A's outdegree.

1.3.1. Geodesic Routes

A geodesic route is the **shortest path** between any two nodes.

- In **Unweighted Networks**, the shortest route between two nodes is defined by which paths has least amount of nodes in-between.
- In **Weighted Networks**, the method above is insufficient. So, instead we use algorithms like Dijkstra's algorithm.
    - If our edge weights show "distance" (or some quantity which we want to minimize) then we apply the algorithm normally.
    - If our edge weights show "bandwidth" (or some quantity which we want to maximize) then we invert the algorithm and then apply it.

1.3.2. Centrality Measures

Centrality is used to find the most important nodes or edges within a network. We can use **Degree** (and also **Strength**) as centrality measures:

> **Degree Centrality** → Picking the center-most node based on the highest degree.

> **Strength Centrality** → Picking the center-most node based on the highest strength, in directed networks.

Degree/Strength Centralities don't work for many networks. For those we need Global Measures like **Closeness** and **Betweenness**:

> **Betweenness Centrality** → Picking the center-most node based on which nodes most paths go through.

Divide how many shortest paths go through a node by how many shortest paths go from every node to every other node to get that node's betweenness. Whichever node has the highest betweenness is the center-most.

> **Closeness Centrality** → Picking the center-most node based on which node is the closest to everyone else.

Calculate the inverse of the sum of all path lengths to get that node's closeness. Whichever node has the highest is the center-most.

1.3. Transitivity

The tendency for networks to form triangles like if A is connected to both B and C, then B and C are also likely to be connected.

> **Global Clustering Coefficient** → The ratio of the number of closed triplets (triangles) over the total number of connected triplets (both closed and open). A closed triplet consists of three nodes all connected to each other.

They defined the triplet values based on four calculation methods:

1. **Arithmetic Mean (A.M):** Uses the average of the link weights.

2. **Geometric Mean (G.M):** Decreases the sensitivity issue associated with the arithmetic mean.
3. **Maximum (Max):** Takes the highest value of the two weights, treating a strong/weak tie triplet equally to one with two strong ties.
4. **Minimum (Min):** Takes the lowest value, treating a strong/weak tie triplet equally to one with two weak ties.

> **Local Clustering Coefficient** $\rightarrow$ The ratio between the number of present links divided by the total number of possible links between a node's neighbors. Its value ranges between 0 and 1.

1.4. Bipartite (Two-Mode) Networks

A two-mode network $\mathrm{G} = (\top, \bot, \mathrm{E})$ consists of two different, **disjoint sets of nodes**: the top nodes ($\top$) and the bottom nodes ($\bot$). Links only exist **between** a node from one set and a node from the other set. There is **no link** between nodes from the same set.

Two-mode networks are mostly not analyzed in their native form due to the lack of available network descriptors and methods designed for this structure.

> **Network Projection** $\rightarrow$ Converting two-mode network into a **one-mode network** to enable analysis. This transformation involves choosing one set of nodes and connecting two nodes from that set if they were both connected to the same node in the other set.

The concept of degree in two-mode networks can be ambiguous.

- **Two-mode degree** $\rightarrow$ The number of nodes in the secondary set that a node in the primary set is connected to.
- **One-mode degree** $\rightarrow$ The standard degree measure observed *after* the network has been projected to a one-mode network.