

## **PART I: DESCRIPTIVE ANALYSIS**

A network (graph) describes a collection of nodes (vertices) and the links (edges) between them. A node can represent anything from individuals or firms or countries, or even collections of such entities. A link between two nodes signifies a direct relation between them. For example, in a social context, a link could be a friendship tie, while in the context of countries, a link may be a free trade agreement or a mutual defense pact. In a directed network where there is a clear distinction between source (the sender of a tie) and target (the receiver of the tie), relationships between two nodes are recorded as either asymmetric, mutual, or null. Asymmetric is when a nomination is unidirectional (i.e. when only one person claims to know/be friends/to have spoken with the other person); a mutual (reciprocal) relation is when both nominate each other; and a null relation is when there is no connection or link that exists.

### **Methods**

In social network analysis, a network can be described through its **centrality** and **connectivity and cohesion** properties.

Steps in this module are as follows:

1. **Centrality Measures** -> to determine relationships between nodes and their position in the network.
2. **Connectivity and Cohesion Measures** -> to examine how tightly connected or clustered the network is.
3. **Evaluation** -> to test the network for small world properties.

## **1 CENTRALITY**

### **1.1 Node Centrality**

Node centrality is a general measure of a nodes “importance” within a network and is often defined in terms of:

**Degree:** Number of nodes a node is connected to (both sending and receiving ties).

**Indegree:** Number of nodes nominating a node (receiving a tie).

**Outdegree:** Number of nodes a node nominates (sending a tie).

**Closeness:** Proximity to each node in the network.

**Betweenness:** Frequency at which this node connects every other node in the network using the shortest path.

**PageRank:** Number of important nodes a node is connected to.

**Eccentricity:** Maximum shortest distance from (or to) a node, to (or from) all other nodes in the graph.

### **1.2 Network Centrality**

Network centrality is a measure of a node’s ties relative to the ties present in the network and the distribution of ties throughout the network. For example, we can determine the extent to which the “importance” of a

node in a network or the “power” of a node in a network is concentrated in a few nodes by examining whether the network’s degree distribution is normally distributed or skewed.

**Degree distribution:** Frequency distribution of degree values of nodes. A skewed degree distribution, where there are a few high degree (popular) nodes and many low degree (periphery) nodes, is evidence of preferential attachment (i.e. the more connected a node is, the more likely it is to make new connections), and therefore concentrated power.

**Density:** Volume of connections in a network. It is the number of ties relative to the number of all possible ties. A density is 0 for a graph without edges and 1 for a complete graph.

**Average Path Length:** Average shortest (geodesic) distance between each starting and ending node (i.e. the average number of steps one has to take across the network for connecting two separate individuals).

## 2 CONNECTIVITY AND COHESION

Connectivity and cohesion properties refer to the direction, frequency and consistency of relations between nodes and the nodes in their neighbourhood [a personal or ego network that only includes nodes a node is connected to]. This includes the study of dyads (relations between 2-nodes), triads (relations between 3-nodes), clusters and cliques (subset of densely connected ties or subgraphs). This can be examined in the following measures:

**Reciprocity:** Ratio of nodes in a nodes neighbourhood that a node is connected to that reciprocate ties.

**Transitivity:** Fraction of all triangles in a nodes neighbourhood where a node is connected to a node that is connected to another node that it is also connected to (e.g a friend of a friend is a friend).

**Hierarchical Agreement:** Number of triads in a nodes neighbourhood where there is a consensus on the directionality of ties (e.g. many subordinates nominating one boss or followers nominating one leader).

The structural cohesion of a network can be defined as the minimum number of actors who, if removed from the network, would disconnect it. One way to measure overall network cohesion is through the *Clustering Coefficient*. This measure offers a sense of the number of routes and paths available in connecting the network.

**Clustering Coefficient:** Extent to which links in a network follow a transitive property (i.e. likelihood of node  $i$  being connected to node  $k$  given that  $i$  is connected to  $j$  and  $j$  is connected to  $k$ ). This captures how tightly knit or cohesive the network is.

A related measure, but not covered in this analysis, describes a network by its similarity to connected node attributes.

**Homophily:** Tendency for nodes with similar attributes to be more likely connected with each other than with nodes of dissimilar attributes.

### 3 EVALUATION

Descriptive measures are not on their own intuitive. As such, a good heuristic to follow when evaluating a network is to assess how much of the current network's features are simulated by a random process, and how closely they resemble a small world. Introduced by Watts and Strogatz (1998), a small world embodies the idea that that unlike random networks of the same size, large networks tend to have a small diameter or small **Average Path Length** and a high **Clustering Coefficient**.

### 4 REFERENCES

Goyal, Sanjeev., *Connections: An Introduction to the Economics of Networks*, Princeton, NJ: Princeton University Press, 2007.

Jackson, Matthew O., *Social and Economic Networks*, Princeton, NJ: Princeton University Press, 2008.

Wasserman, Stanley and Katherine Faust, *Social Network Analysis*, New York, NY: Cambridge University Press, 2007.

Watts, Duncan J., *Small Worlds: the dynamics of networks between order and randomness*, Princeton, NJ: Princeton University Press, 2005.

Copeland, Molly., "Whole Network Descriptive Statistics," *Social Networks and Health Workshop*, 2019. Available at: <https://sites.duke.edu/dnac/07-whole-network-descriptive-statistics/>