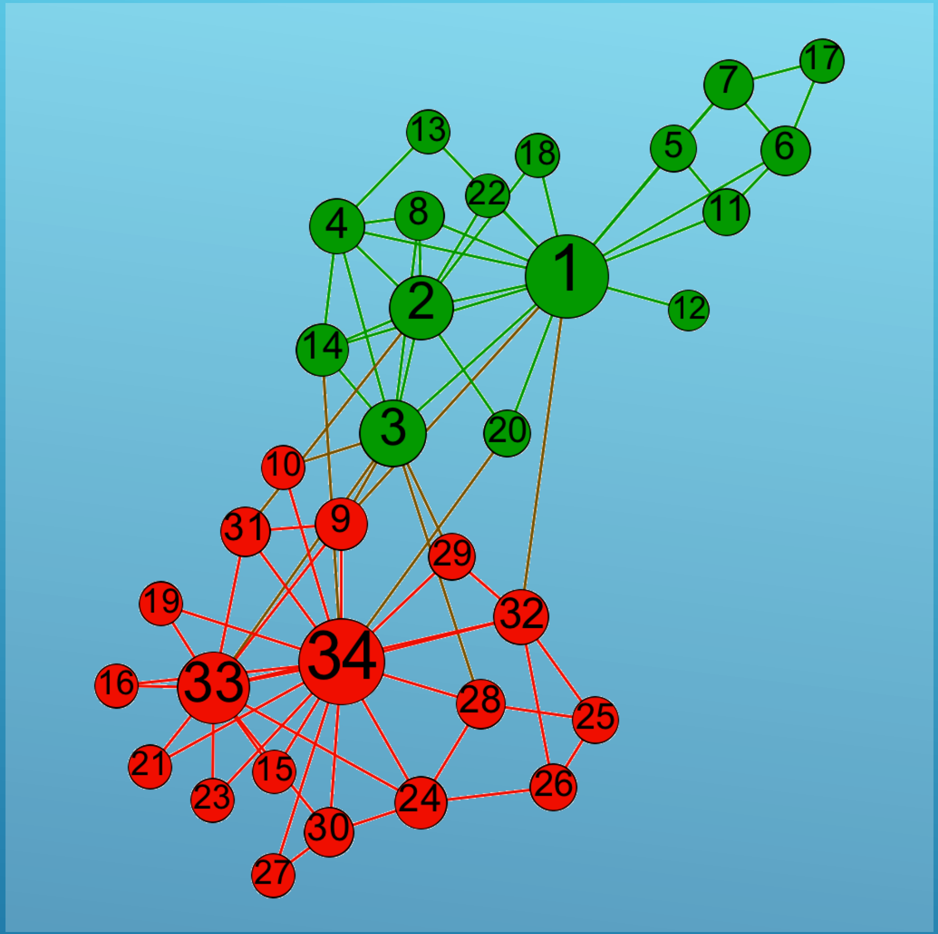
BASIC CONCEPTS

Analysis of Large Scale Social Networks

Bart Thijs



Karate Club

34 Nodes, 78 Edges
Node 1 = Instructor
Node 34 = President

Edges refer to interactions between members

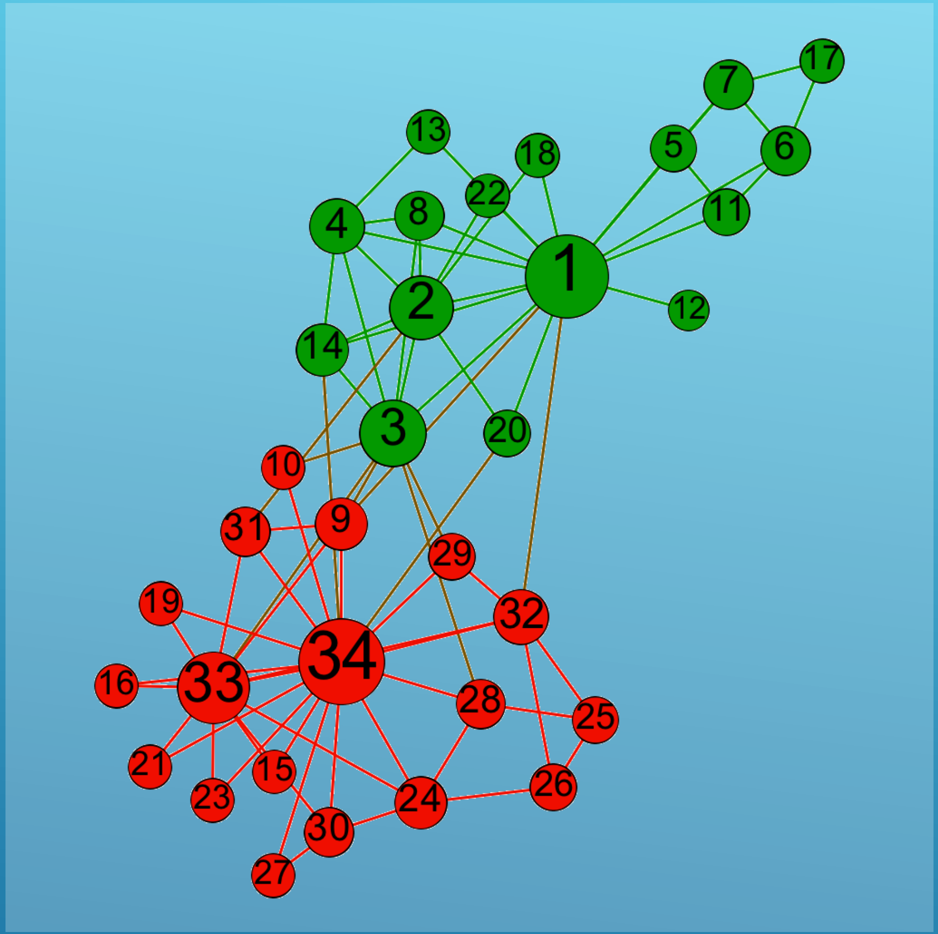
- Unweighted and Undirected

Formal:

A graph is an ordered pair of two sets
$G = (V, E)$
V = set of vertices
E = set of edges which are pairs of elements from V

NETWORK – NODES AND EDGES

Zachary, A, 1977, An Information Flow Model for Conflict and Fission in Small Groups, *Journal of Anthropological Research*, 33 (4), 452-473.

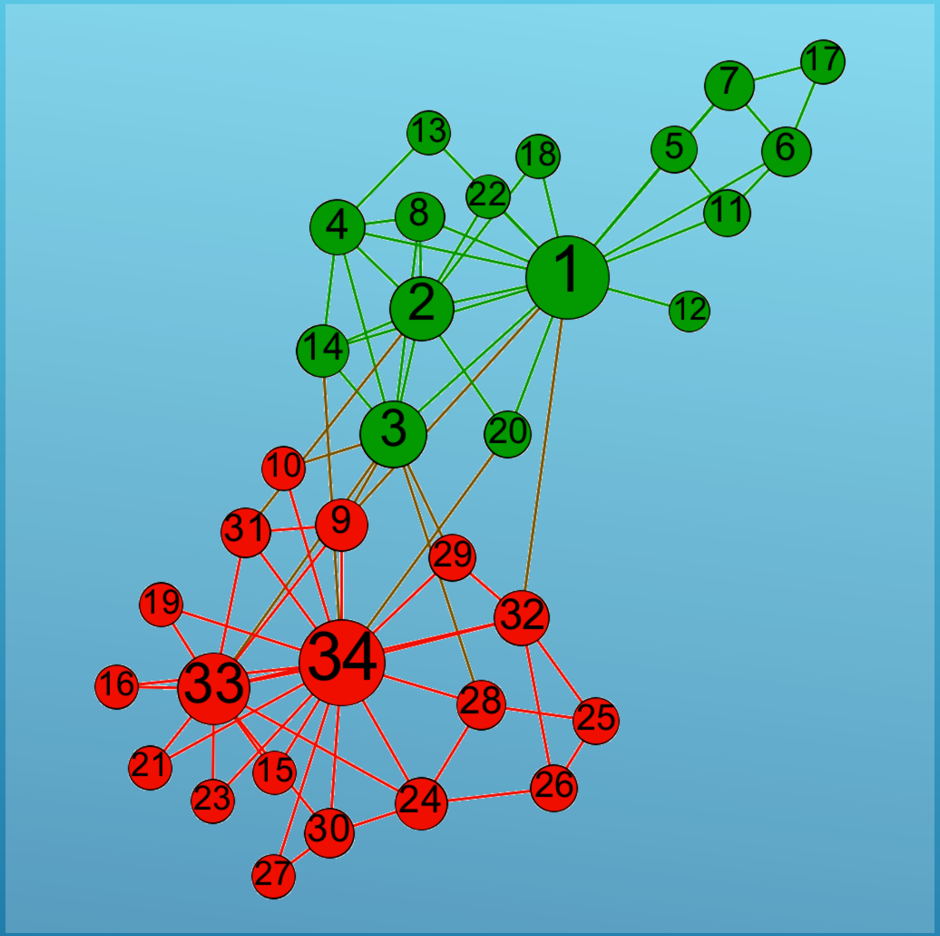


Properties at two distinct levels:

- Local level
  This discusses the role or position of individual nodes or small subgraphs within the network. Eg. The role of node 20 in the karate club: He has a limited number of links but has contacts with both the trainer and the administror

- Global level
  This type of analysis focusses on the large scale structure of the network or large subgraphs.

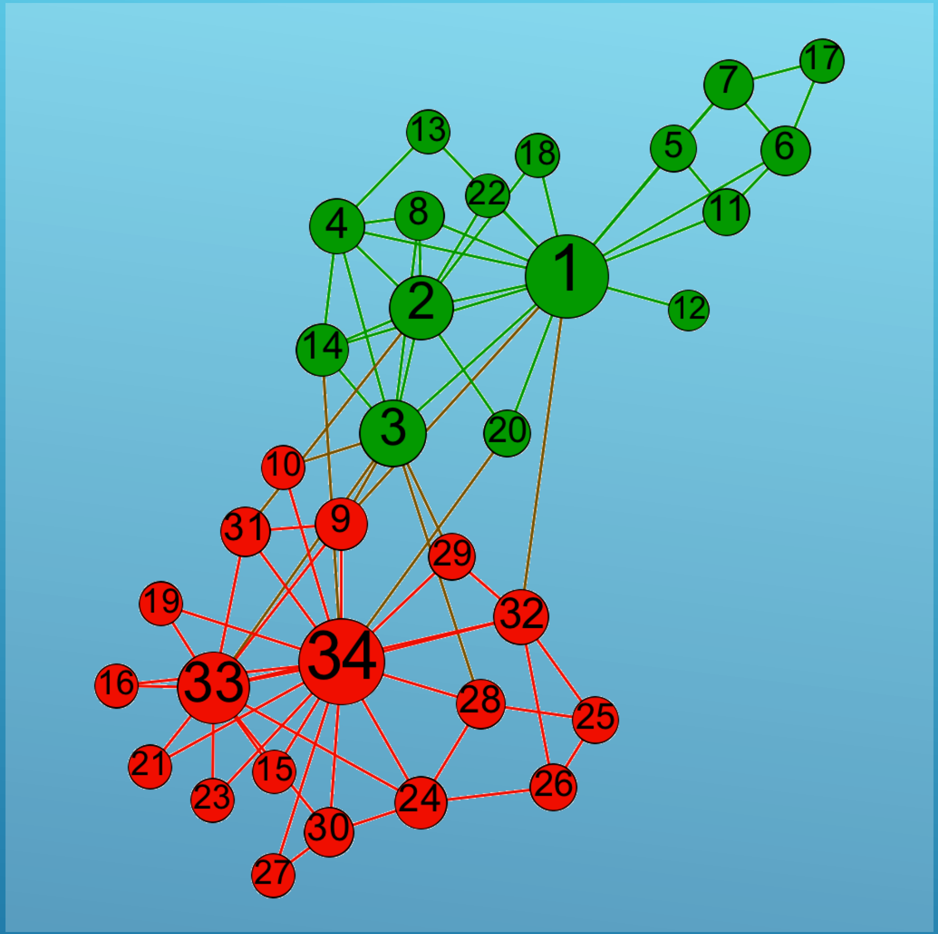
LOCAL AND GLOBAL NETWORK ANALYSIS



Centrality

- Degree
  Number of links associated with a node
- Farness / Closeness
  Farness is the sum of distances between a node and all the other nodes. Closeness is the inverse
- Betweenness
  This quantifies the times that a node acts as a bridge between pairs of other nodes
- Pagerank
  Nodes get a high score if they are connected with other high scoring nodes.

LOCAL LEVEL CENTRALITY



Degree of a node:

Number of links to other nodes

In-Degree = number of incoming links

Out-Degree = outgoing

In directed networks

$k = k^{in} + k^{out}$

$$L = \sum_{i=1}^{N} k_i^{in} = \sum_{i=1}^{N} k_i^{out}$$

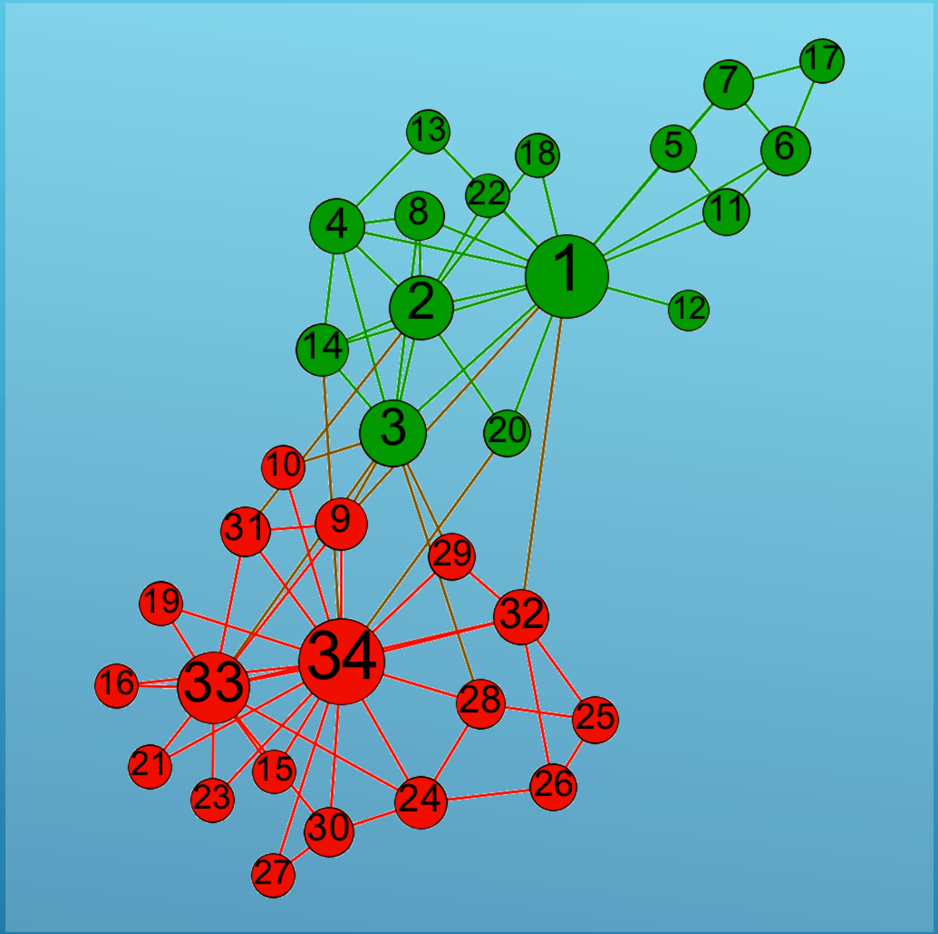
N = Number of nodes

L = Number of edges

k_i = Degree of node i

$$L = \frac{1}{2} \sum_{i=1}^{N} k_i$$

LOCAL LEVEL: DEGREE

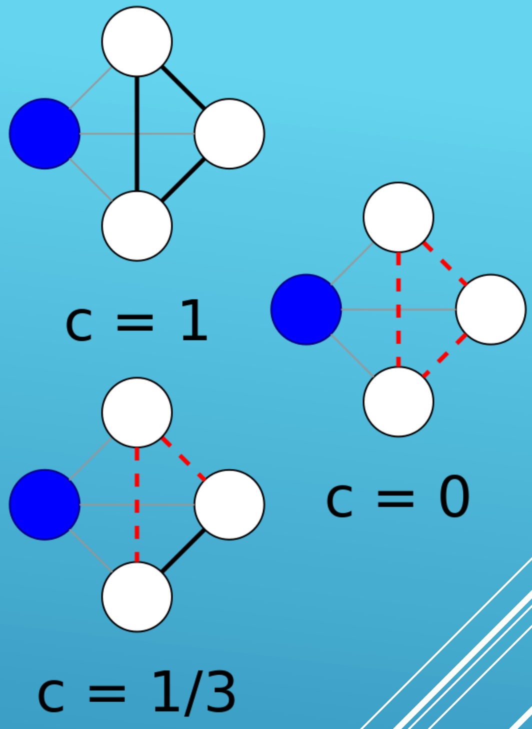


Neighbourhood

The neighbours of a node are those other nodes that are connected with the source.

Clustering of a node

The proportion of actual edges between the neighbours of a node and the number of possible links.



Source: Wikipedia

LOCAL LEVEL: CLUSTERING

- Complete network or fully connected.
  All possible links are present in the network

$$L_{max} = \frac{N(N-1)}{2}$$

- Density (Global clustering coefficient)

$$D = \frac{2L}{N(N-1)} = \frac{L}{L_{max}}$$

In most real world networks

$$L \ll L_{max}$$

GLOBAL LEVEL DESCRIPTIVE STATISTICS

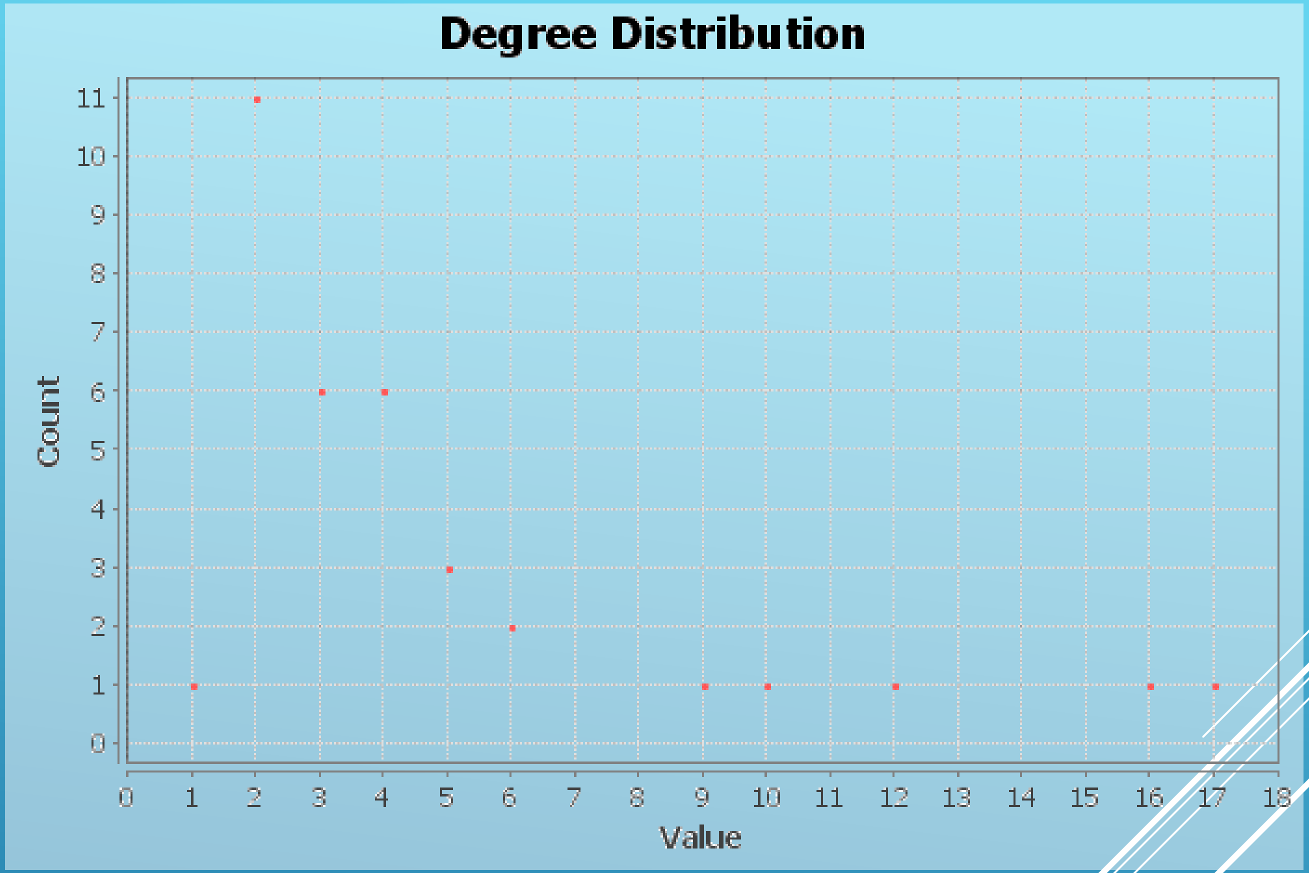
- Average degree

$$\frac{\sum_{i=1}^{N} k_i}{N} = \frac{2L}{N} = D(N-1)$$

- Degree distribution
  Probability distribution

$$\mathrm{P(k)} = \frac{N_k}{N}$$

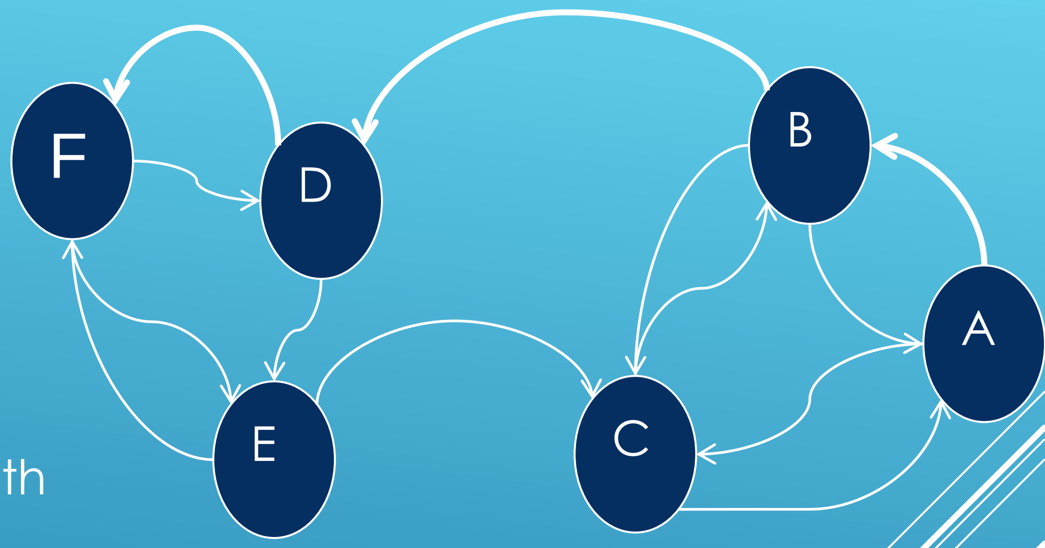
These distributions are often highly skewed with many nodes with low degree and a few nodes with high degree.



GLOBAL LEVEL: DEGREE

- A *Walk* is any sequence of nodes and edges, starting and ending with a node. Nodes and edges can occur more than once. The length is the number of edges in the sequence.
- A *Trail* is a walk in which an edge can only occur once.
- A *Path* is a walk in which both nodes and edges only occur once. The length is the number of edges

- A *Shortest Path* between two nodes i and j is the path with the least edges, with the smallest length. This is also called *geodesic*
- *Network Diameter* is the length of the longest shortest path present in a graph
- *Average Path length* is the average of all shortest paths between any two nodes



GLOBAL LEVEL: TRAVERSALS

GLOBAL LEVEL: COMPONENTS

- Connected vs Disconnected
  A graph is connected if it is possible to find a path between every pair of nodes within the graph. A graph is disconnected if there exists at least one pair of nodes where it is not possible to reach each other through traversal.
- Connected component
  This is a special subset of a graph or subgraph in which it is possible to draw a path between any pair of two nodes in the component and it is impossible to draw a path between any pair of nodes with one in the component and one outside of it.
- Largest connected component
  This the component containing most vertices within the network
- Singleton
  A single node which has no connection to any of the other nodes in the network.

MATRIX REPRESENTATION OF A NETWORK

Adjacency Matrix

$$A : N\ by\ N$$

A_{ij} = 1 if link from j to i , otherwise 0

In undirected network

$A_{ij} = A_{ji}$

$$k_i = \sum_{j=1}^{N} A_{ji} = \sum_{j=1}^{N} A_{ij}$$

In directed network

$$k_i^{in} = \sum_{j=1}^{N} A_{ij}\,, \qquad k_i^{out} = \sum_{j=1}^{N} A_{ji}$$

0	1	1	1	1	1	1	1	1	0	1	1
1	0	1	1	0	0	0	1	0	0	0	0
1	1	0	1	0	0	0	1	1	1	0	0
1	1	1	0	0	0	0	1	0	0	0	0
1	0	0	0	0	0	1	0	0	0	1	0
1	0	0	0	0	0	1	0	0	0	1	0
1	0	0	0	1	1	0	0	0	0	0	0
1	1	1	1	0	0	0	0	0	0	0	0
1	0	1	0	0	0	0	0	0	0	0	0
0	0	1	0	0	0	0	0	0	0	0	0
1	0	0	0	1	1	0	0	0	0	0	0
1	0	0	0	0	0	0	0	0	0	0	0

In a weighted network, the links can have a value

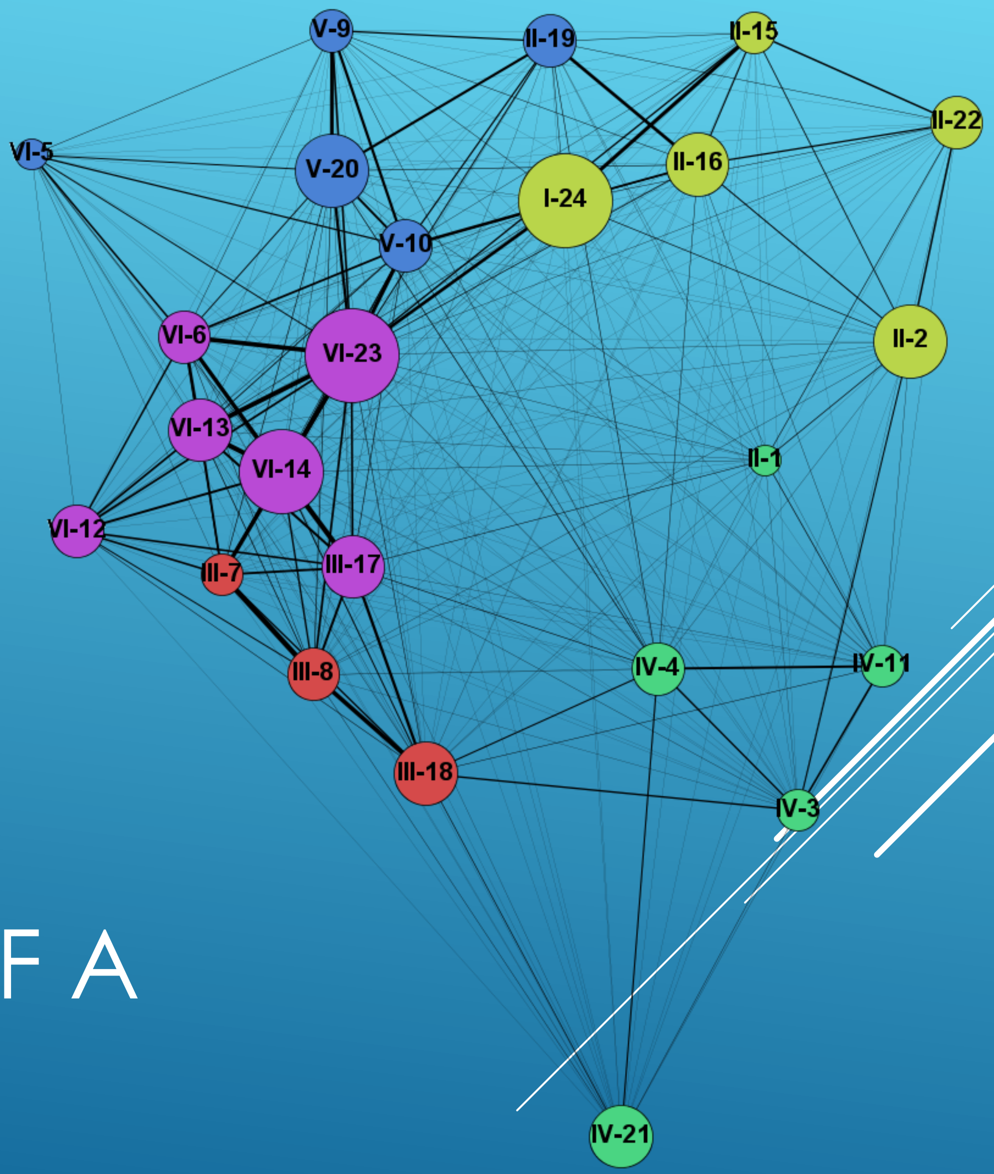
$$A_{ij} = w_{ij}$$

Values can represent:

- Similarities
  Higher weight => Stonger tie between nodes.
  Weighted degree:

$$k_i^w = \sum_{j=1}^{N} w_{ij}$$

- Distances
  Higher weight => nodes are further away from each other, a weaker tie



MATRIX REPRESENTATION OF A WEIGHTED NETWORK