Network Analysis

Dr. Shaina Race
Institute for Advanced Analytics

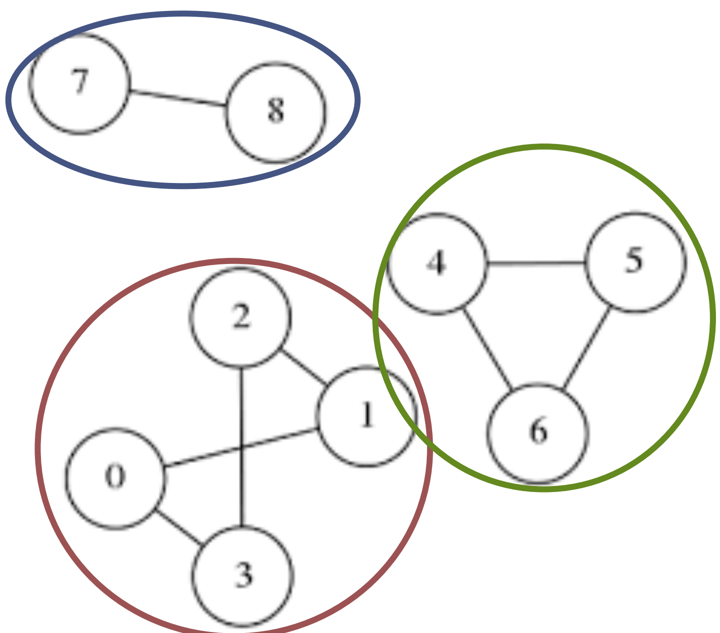
Spring 2018

Descriptives of Network Structure

Components, Cliques, Bridges, Brokers

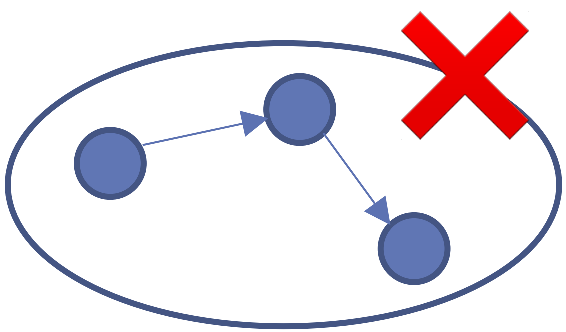
Connected Components

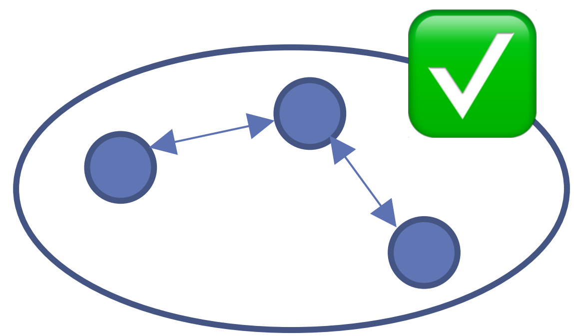
- A graph is **connected** if every node can be reached from every other node
  - (no separate pieces)
- A **component** of a graph is a collection of nodes which are connected themselves but disconnected from the rest of the graph.



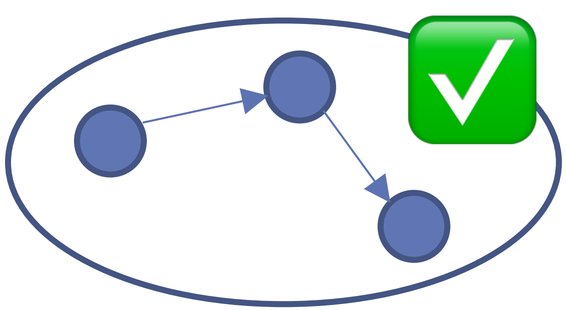
Connected Components (in Directed Networks)

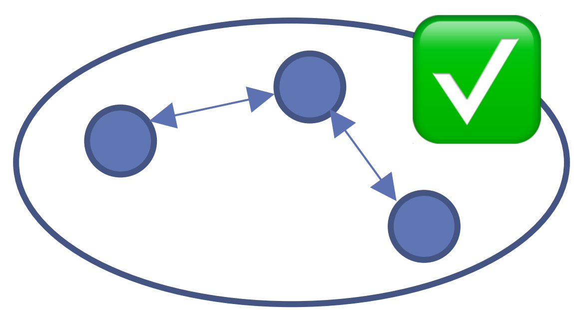
- **Strongly Connected**: All nodes must be connected by directed path in both directions





- **Weakly Connected**: Nodes connected by edges regardless of direction

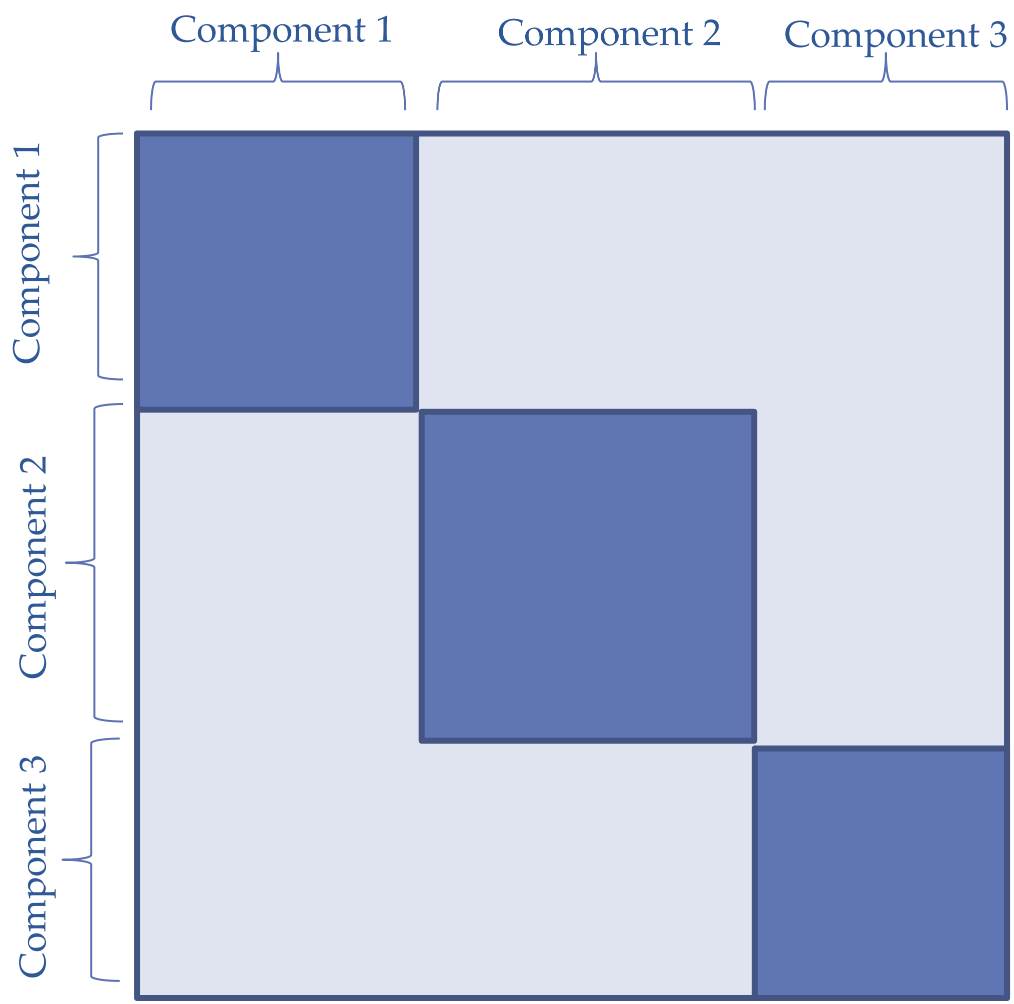




Pop Quiz

If a network has more than one connected component, what does that say about it's adjacency matrix?

Solution



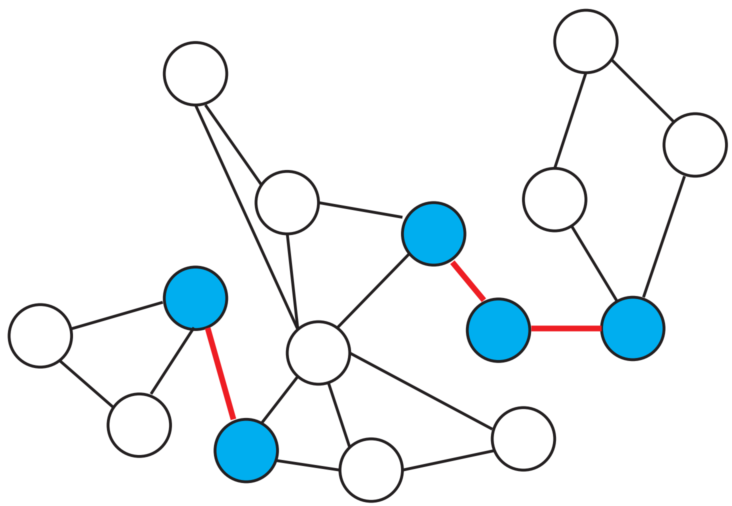
It’s ”block diagonal”

Utility of Connected Components

- In retail setting, find *families* or other purchasing units according to shared traits:
  - Form network with edges between individuals if they share an email or a credit card or a license plate etc.
  - The connected components of this network could create family IDs
- In fraud setting, similar analysis might provide fraudulent networks of claims.

Bridges and Brokers

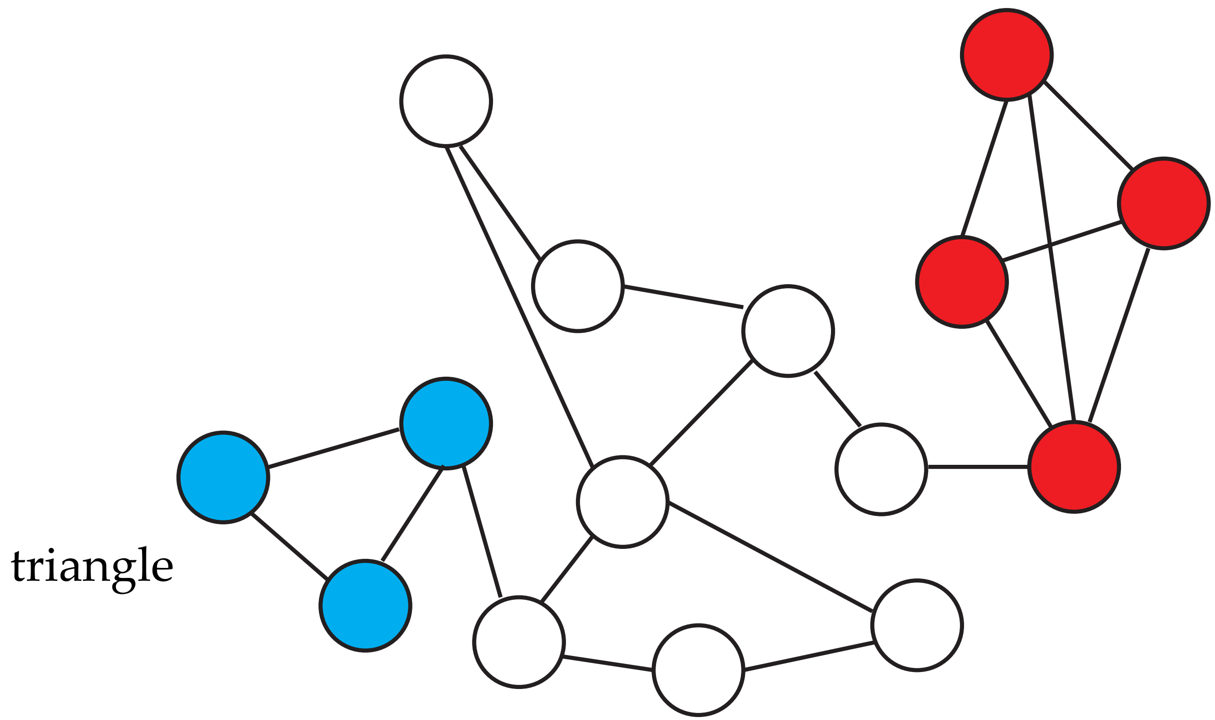
- A **bridge** is an edge whose removal disconnects the network.
- A **broker** is a node whose removal disconnects the network.



- Important players in the “small world effect.”

Cliques
(aka Complete Graphs)

- A **clique** is a group of *three or more* nodes among which all possible edges exist. Each node in a clique is connected to every other node in that clique.



N-Cliques

- Large cliques hard to find. Relax the definition to a group of nodes separated by a distance ≤ n

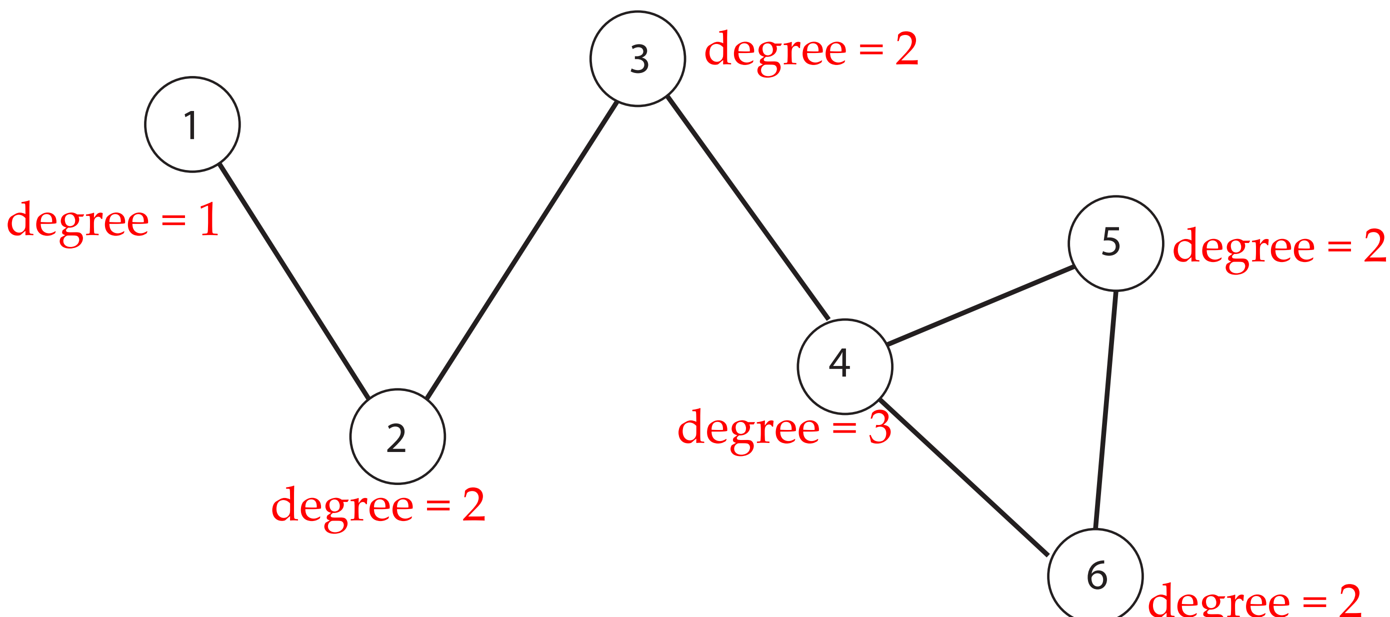
- 2-clique: group of nodes such that each is either directly connected to or shares at least one neighbor with every other node.

Nodal Degree

weighted/unweighted, directed/undirected
and distribution across a network.

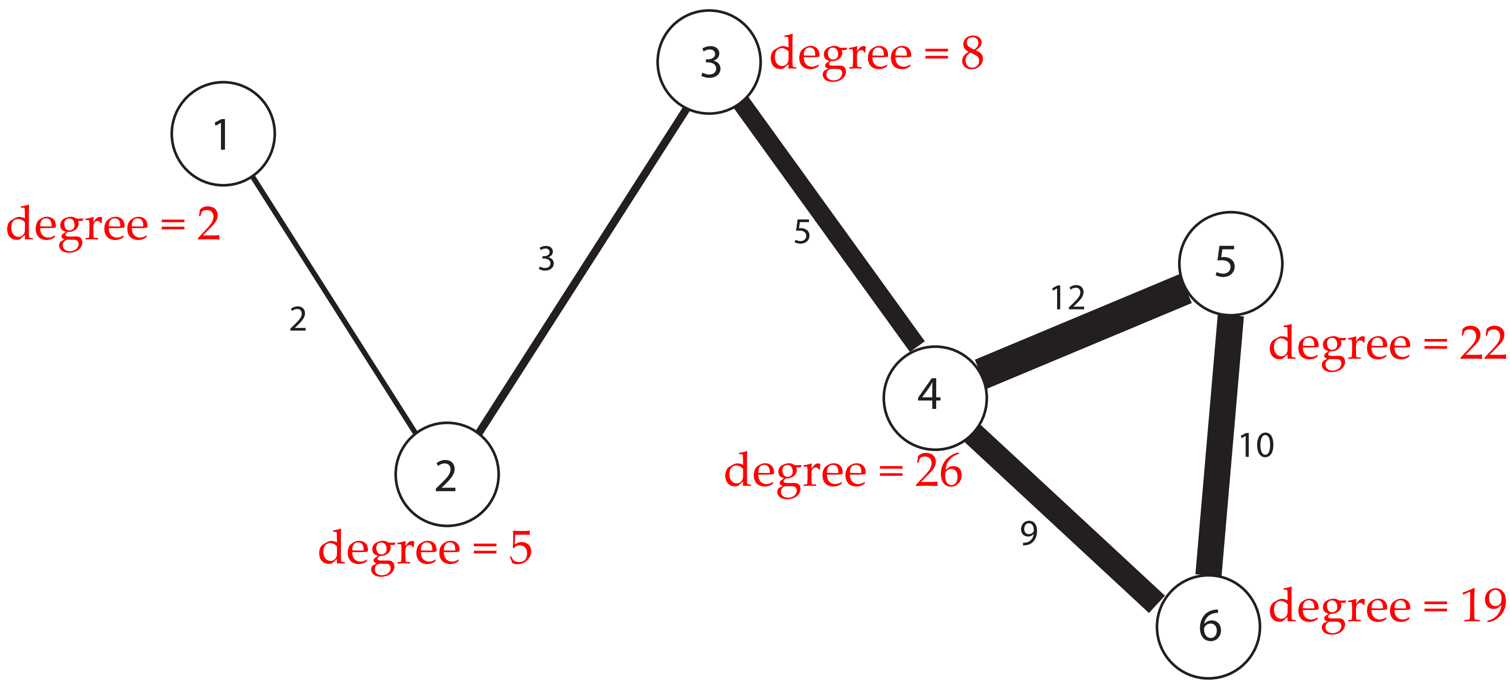
Degree

- The degree of a node measures the connectedness of that node in the network.
- For a binary graph, it is simply the number of edges connected to that node.



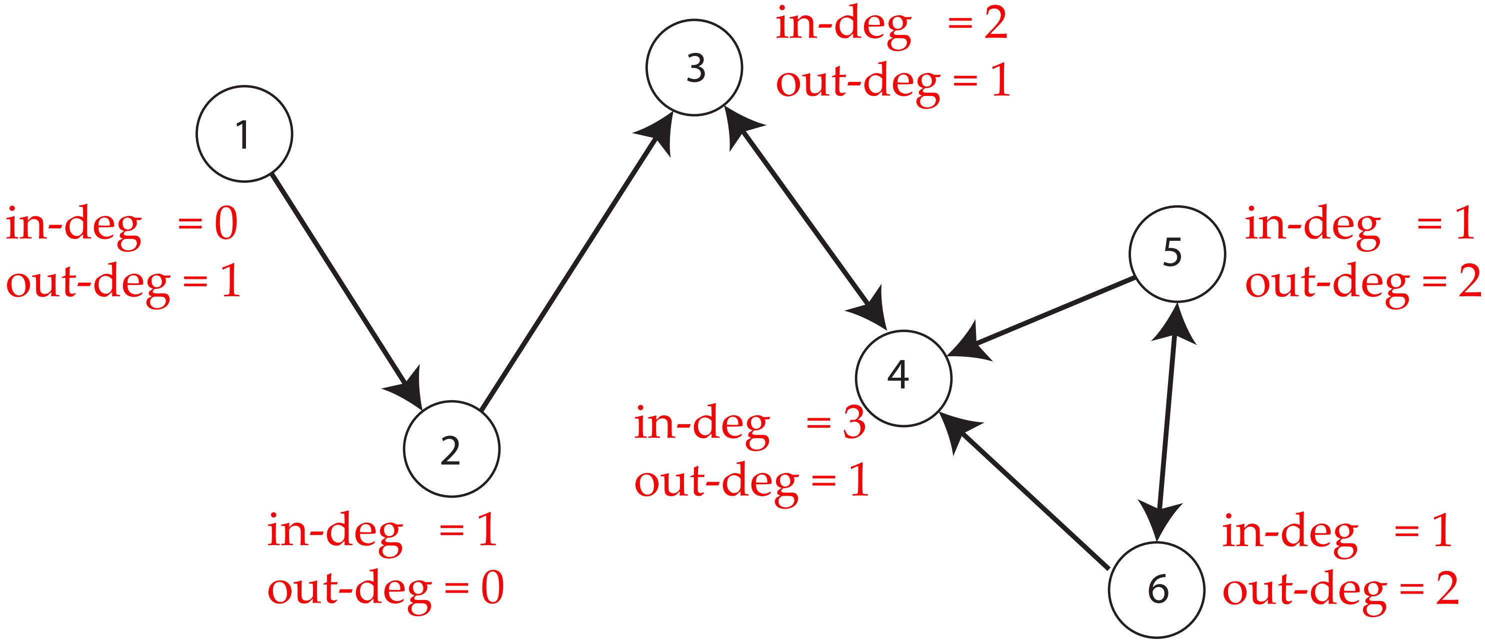
Degree

- For a weighted graph, it is the sum of the weights of edges connected to that node.



Degree

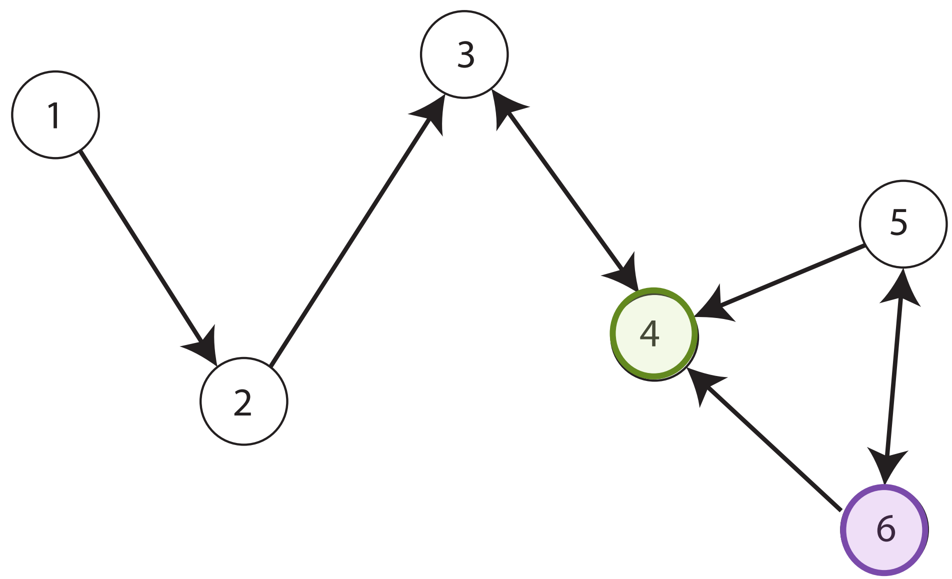
- For a directed graph, we calculate both an in-degree and an out-degree.



Degree

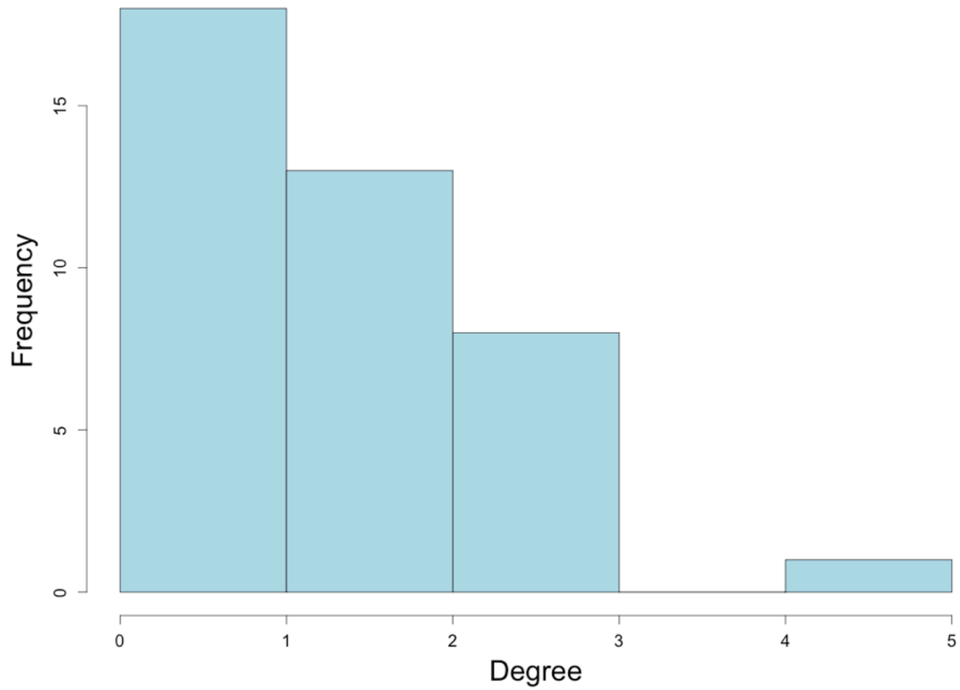
- Nodal degrees are sums of rows and/or columns of the corresponding adjacency matrix.

$$
\begin{matrix}
0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 1 & 1 \\
0 & 0 & 0 & 1 & 0 & 1 \\
0 & 0 & 0 & 1 & 1 & 0
\end{matrix}
$$



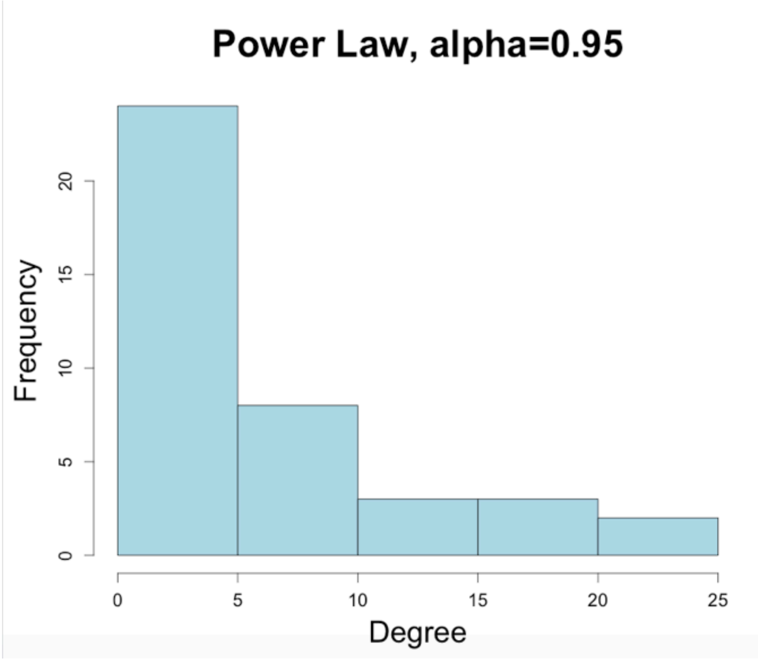
Degree Distribution

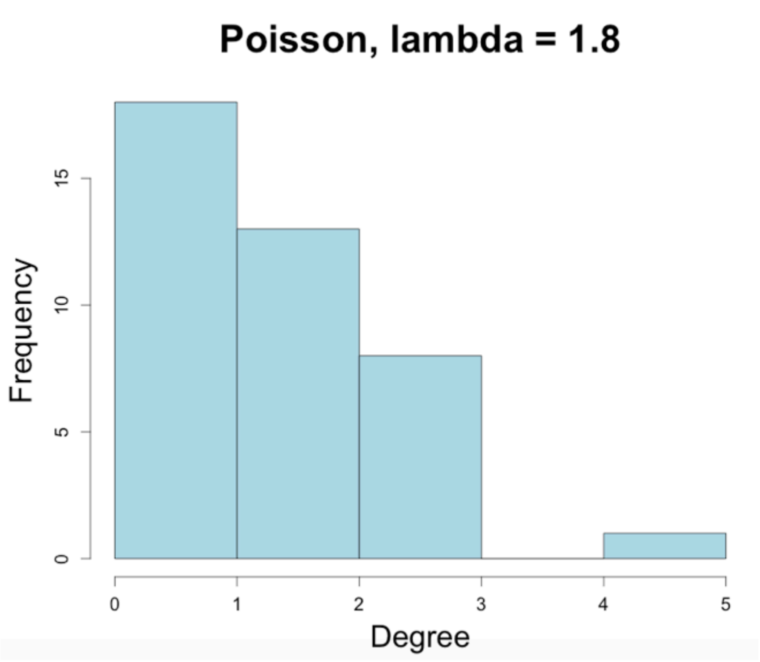
- It's common to look at the distribution of degrees in a network.
- Usually many nodes with low degree and few with high degree.



Degree Distributions

- It's known that most networks follow natural patterns when it comes to degree distribution.
- Two most common distributions:



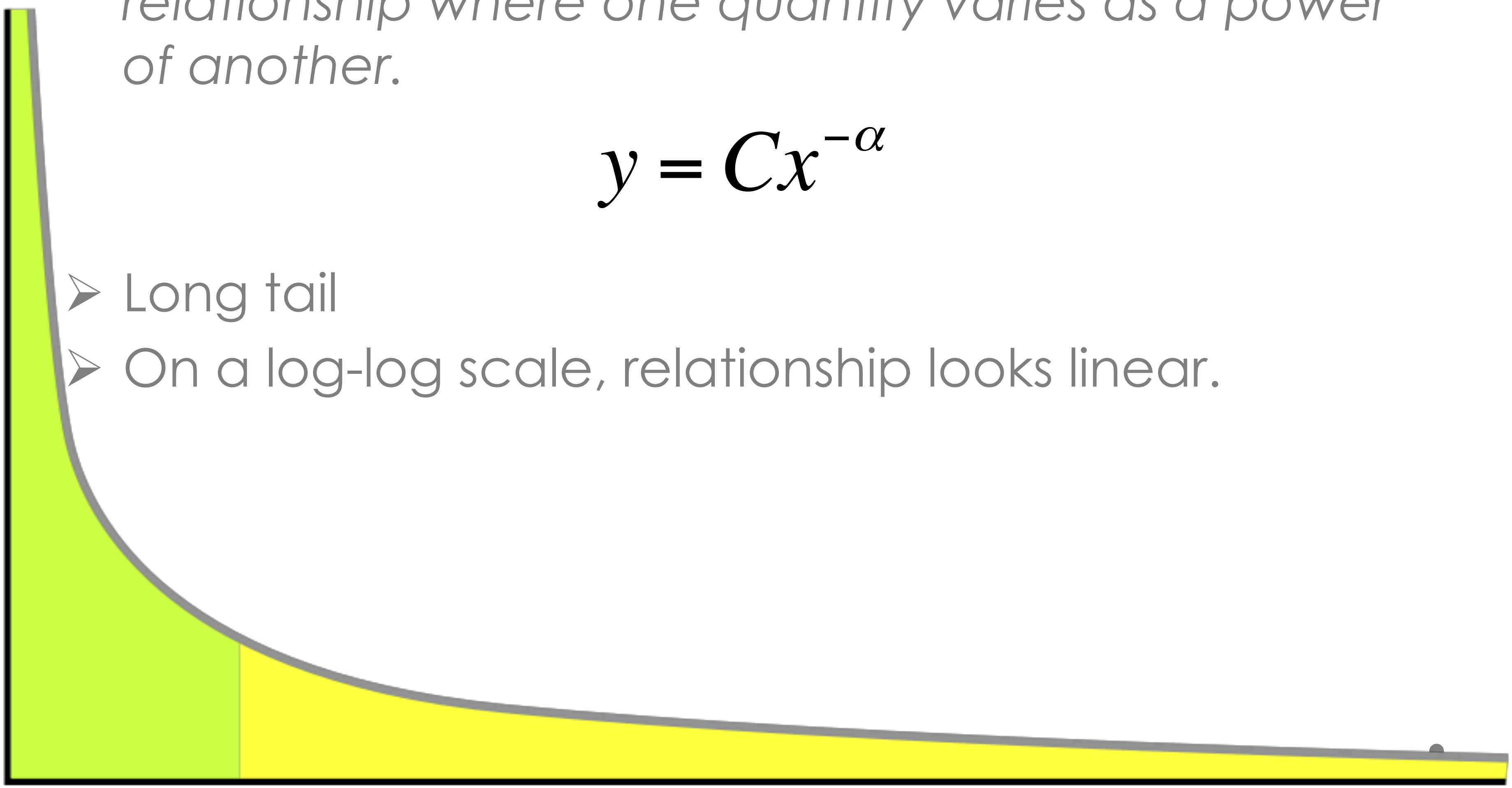


Power Law

- The degree distribution appears as a ***power law**: a relationship where one quantity varies as a power of another.*

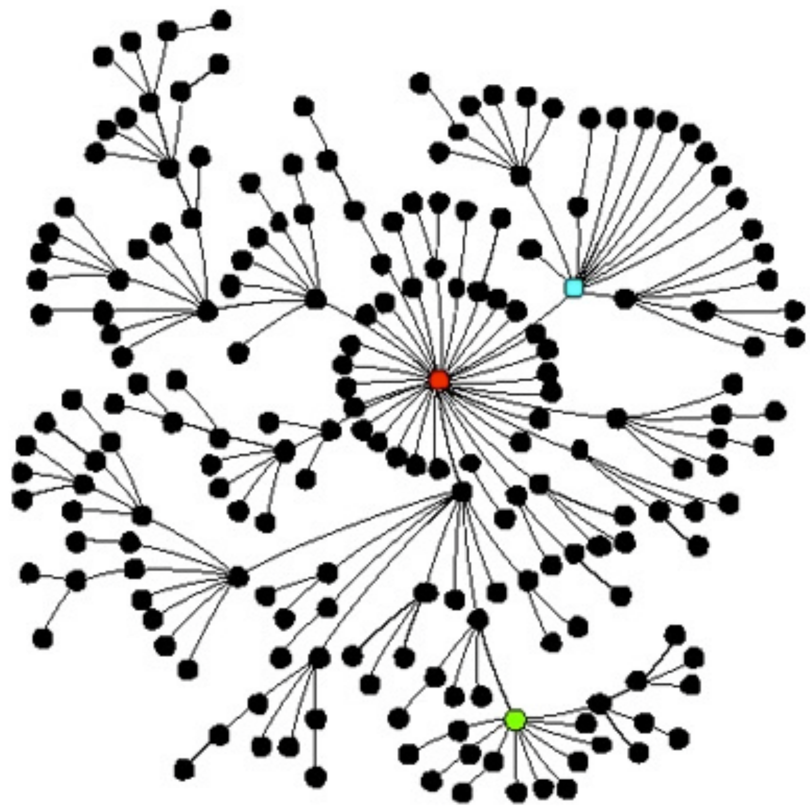
$$y = Cx^{-\alpha}$$

- Long tail
- On a log-log scale, relationship looks linear.



Power Law Graphs
aka Scale Free Networks

- Power law graphs contain a few **hubs** (highly connected nodes) but the majority of nodes in the network have low degree.



Power Law Graphs aka Scale Free Networks

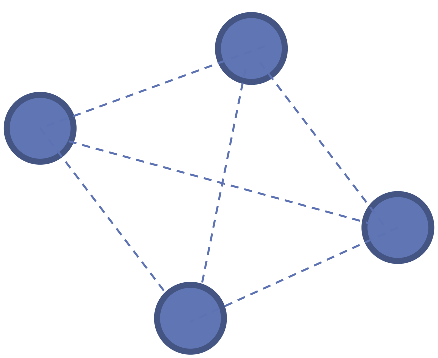
- Properties
  - Robust to random breakdown
  - Vulnerable to targeted attacks
  - Viruses can persist even at low transmission rates
- Real World Examples
  - Email Networks
  - World Wide Web
  - Intranets
  - Diseases with short transmission window
  - Needle Sharing
  - Sexual Contacts

Other Descriptives

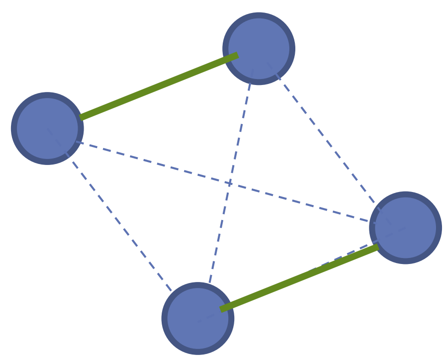
Density, Shortest Paths, Eccentricity, Clustering Coefficients

Density of Graph

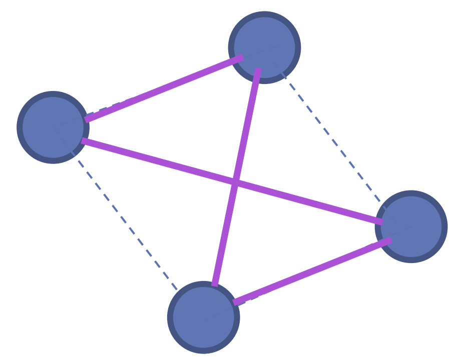
- The **density** of a graph measures how interconnected the nodes are.
- Simply the proportion of possible edges that actually exist in the graph.



6 possible edges



Density = 2/6 = 33%



Density = 4/6 = 66%

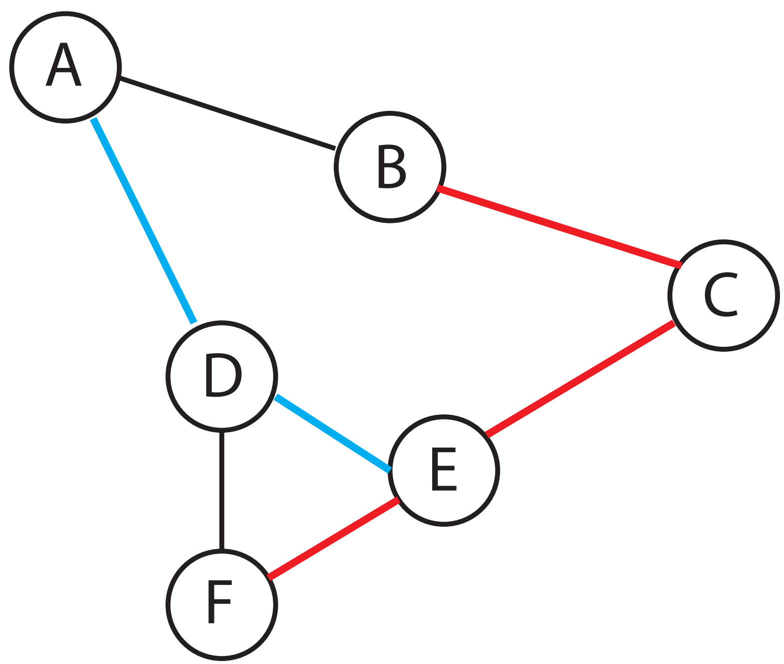
Density of Graph

- Let E be the number of edges in the graph
- Let N the number of vertices.
- The density, Δ, is then:
  - $$\Delta = \frac{2E}{N(N-1)}$$ For undirected graphs
  - $$\Delta = \frac{E}{N(N-1)}$$ For directed graphs

Shortest Paths

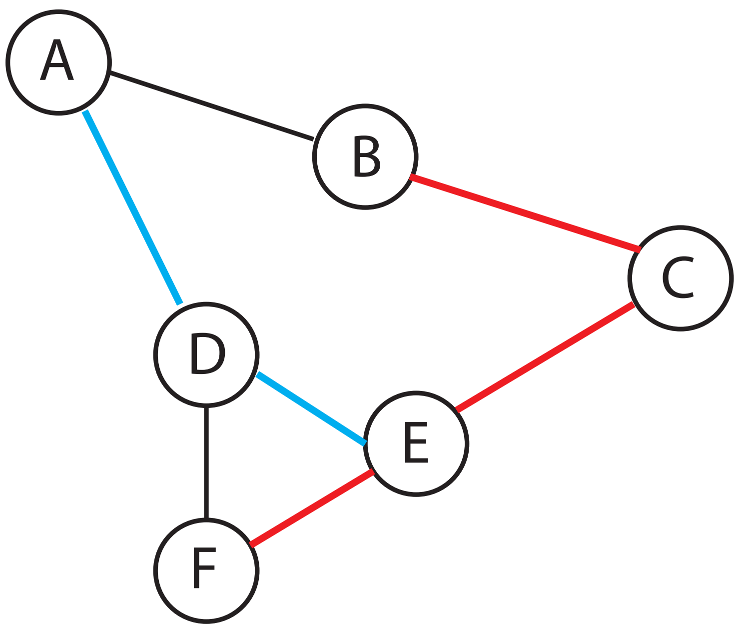
(Geodesic Distances)

- The **geodesic or graph distance** between two vertices is the length of the shortest path from one vertex to the other.
- For directed graph, must be a directed path



Diameter/Eccentricity

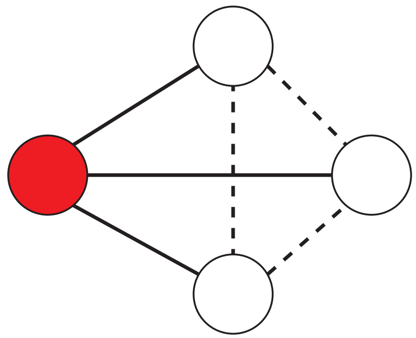
- **Graph Diameter**: Largest Geodesic Distance in the whole network
- **Eccentricity of a node:** Distance to furthest node from that node.



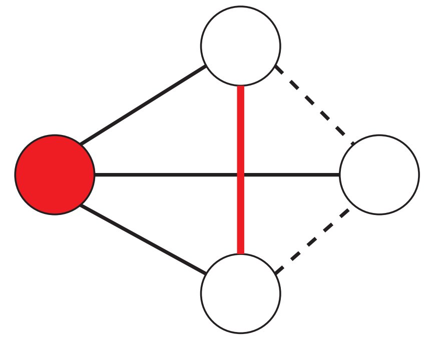
Clustering Coefficient

- The **clustering coefficient** of a node is a measure of the extent to which its neighbors are also neighbors of each other.
- Measures **Transitivity**: if A is connected to B and B is connected to C what is the probability that A is connected to C?
- Ratio of number edges existing between neighbors to those that could possibly exist.

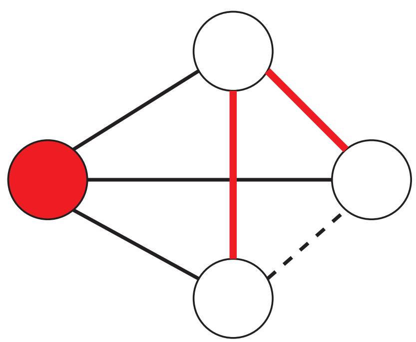
Clustering Coefficient



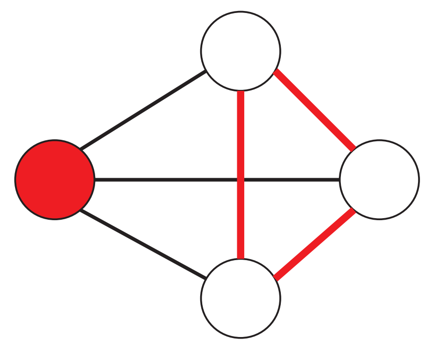
c=0



c=1/3



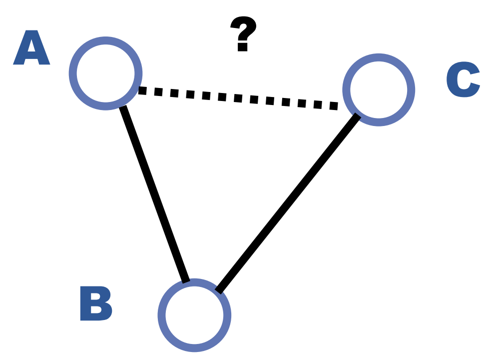
c=2/3



c=1

Clustering Coefficients for Entire Network

- Measure the transitivity of the entire network – does the transitive property hold most of the time?



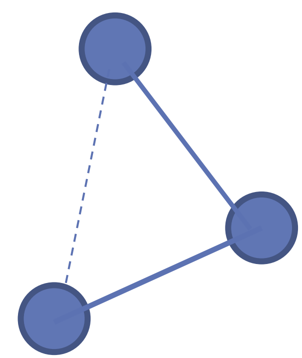
Clustering Coefficients for Entire Network

- **Network Average Clustering Coefficient**: Simply average the clustering coefficient for each node.

- **Global Clustering Coefficient:** Proportion of connected triples that make triangles

$$C = \frac{3 \cdot number\ of\ triangles\ in\ graph}{number\ of\ connected\ triples\ of\ vertices}$$

  - Connected triple is 3 vertices joined by 2 edges.
  - Each triangle makes 3 connected triples.



Fun Links

- Analyze your LinkedIn Network   www.socilab.com
- www.theyrule.net