Network Analysis

Dr. Shaina Race
Institute for Advanced Analytics

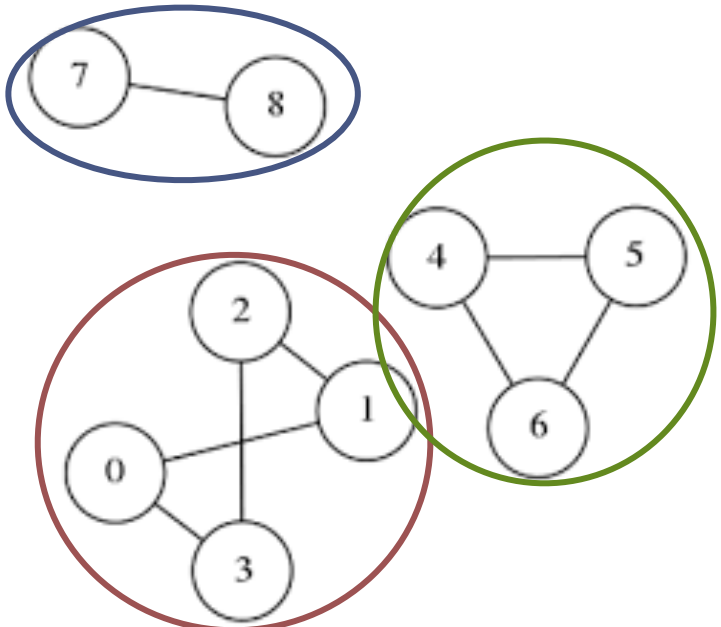
Descriptives of Network Structure

• • •

Components, Cliques, Bridges, Brokers

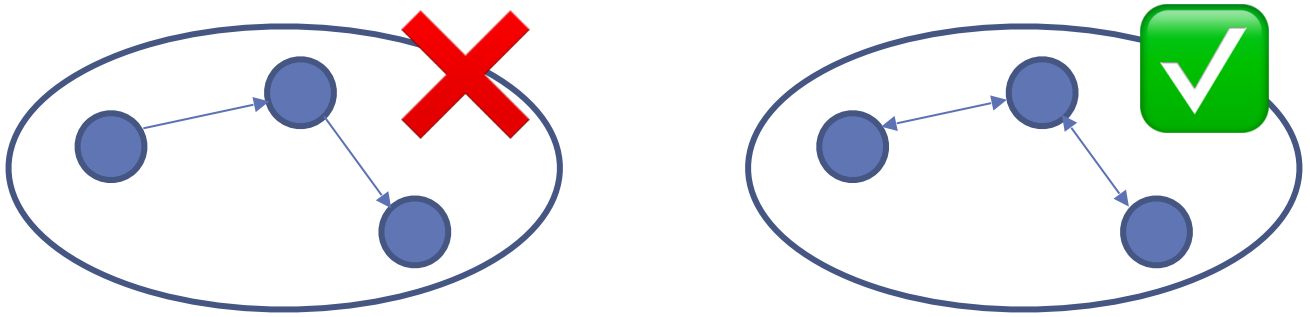
Connected Components

- 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.

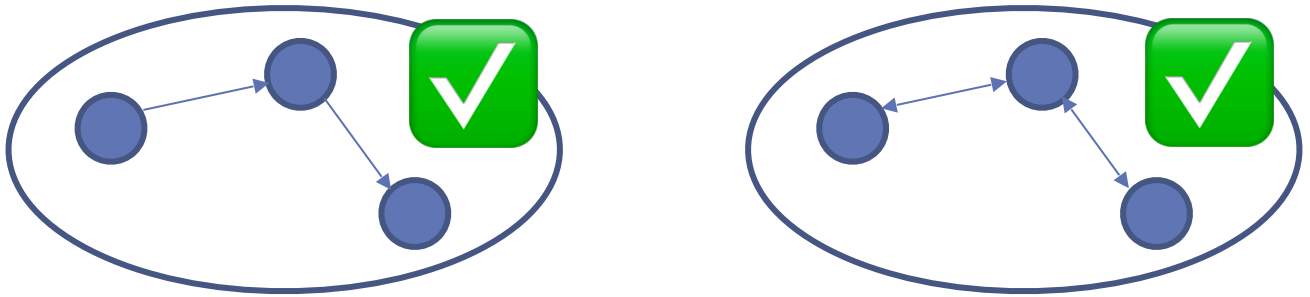


Connected Components (in Directed Networks)

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



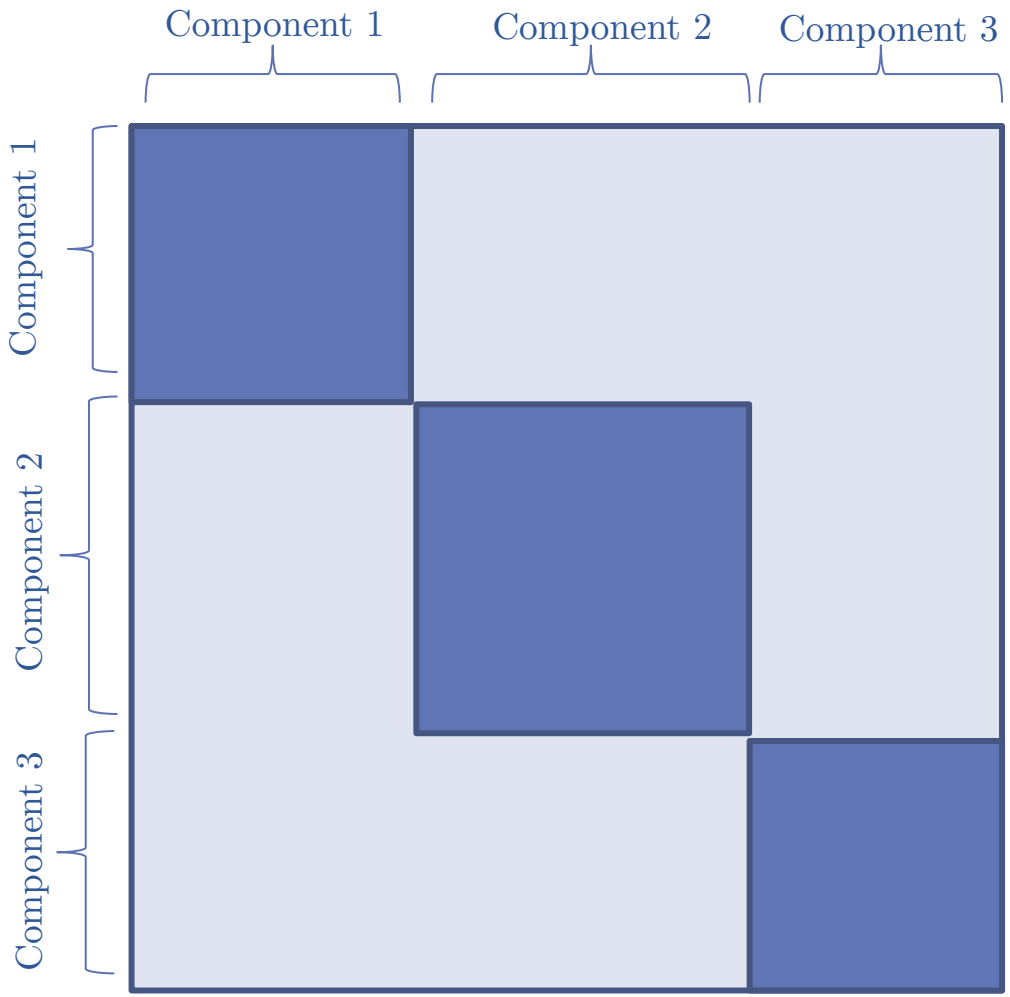
- **Weakly Connected:** Nodes connected by edges regardless of direction



Pop Quiz

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

Solution



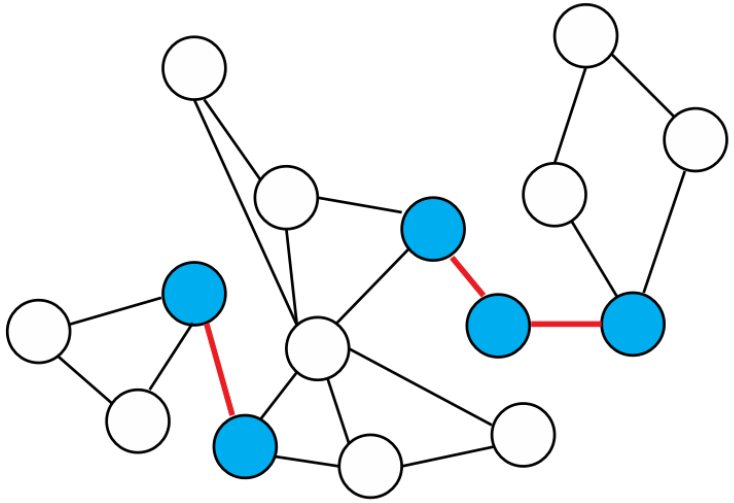
It's "block diagonal"

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

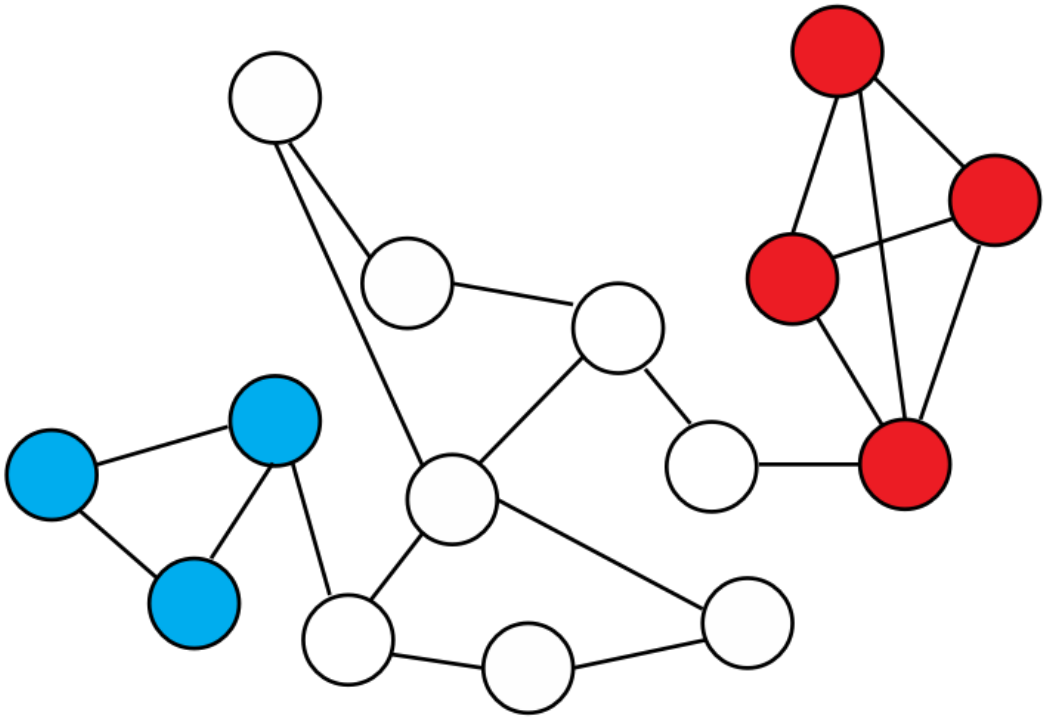
- A **bridge** is an edge whose removal disconnects the network.
- A **broker** is a node whose removal disconnects the network.



- 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.



triangle

Nodal Degree

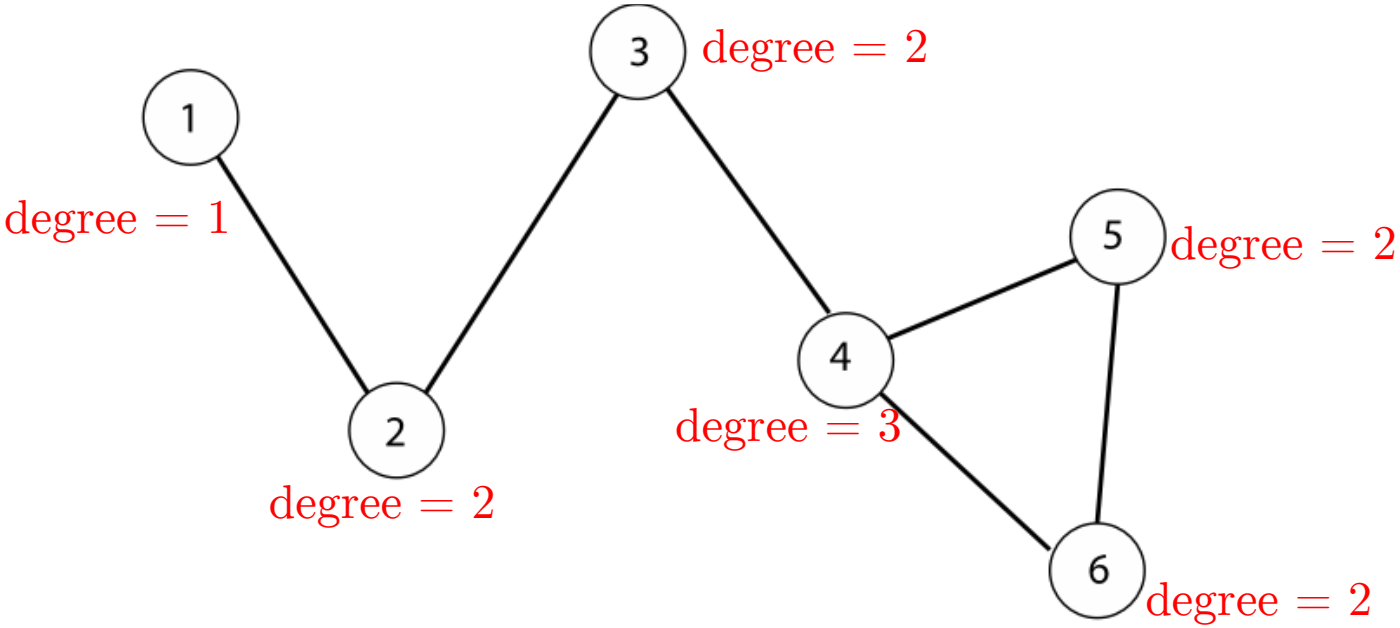
• • •

weighted/unweighted, directed/undirected

and distribution across a network.

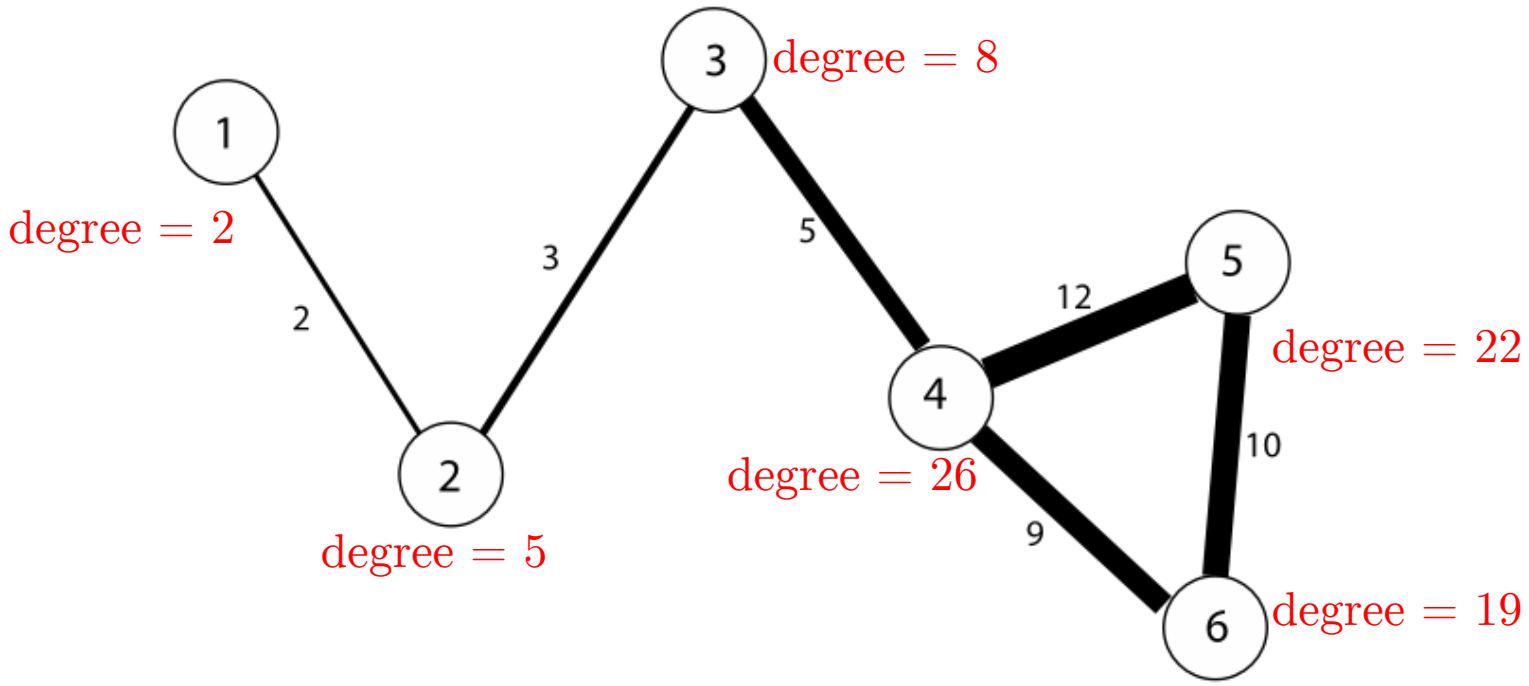
Degree

- 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.



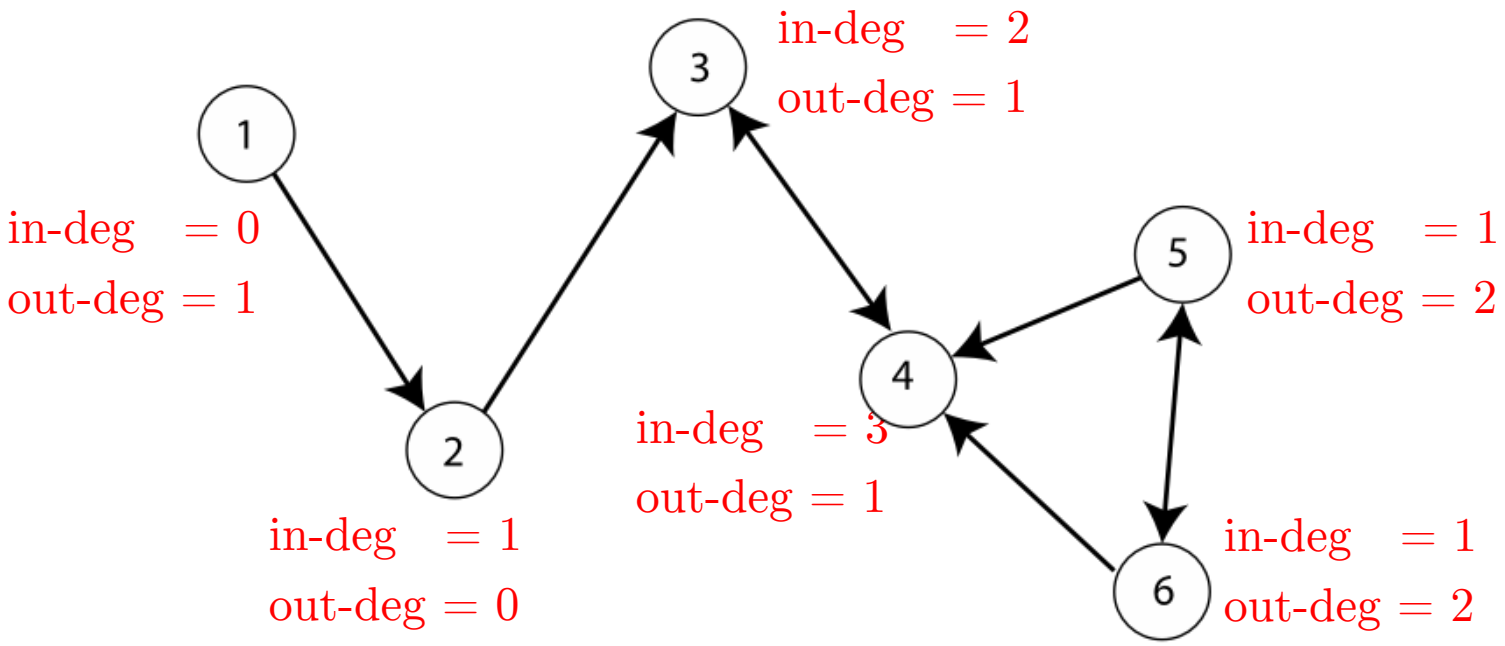
Degree

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



Degree

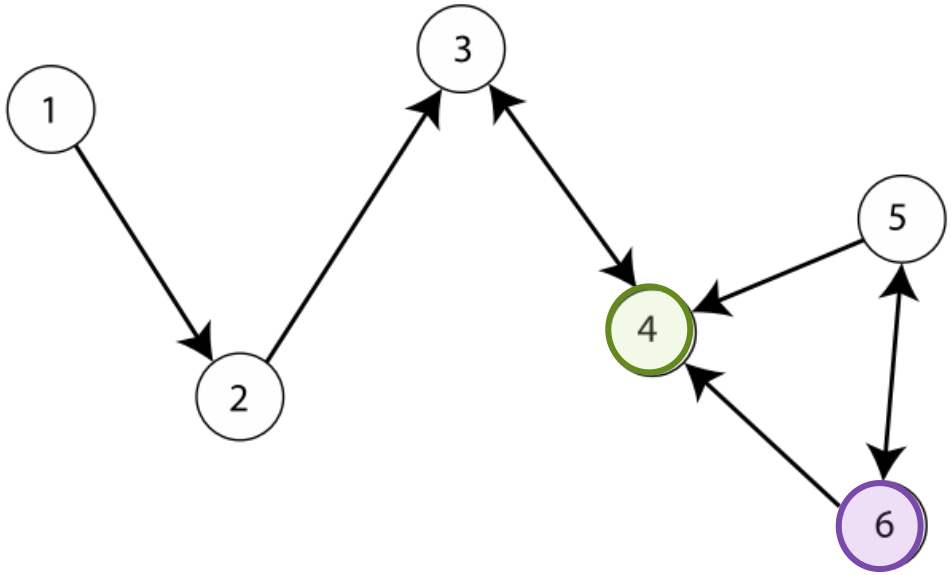
For a directed graph, we calculate both an in-degree and an out-degree.



Degree

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

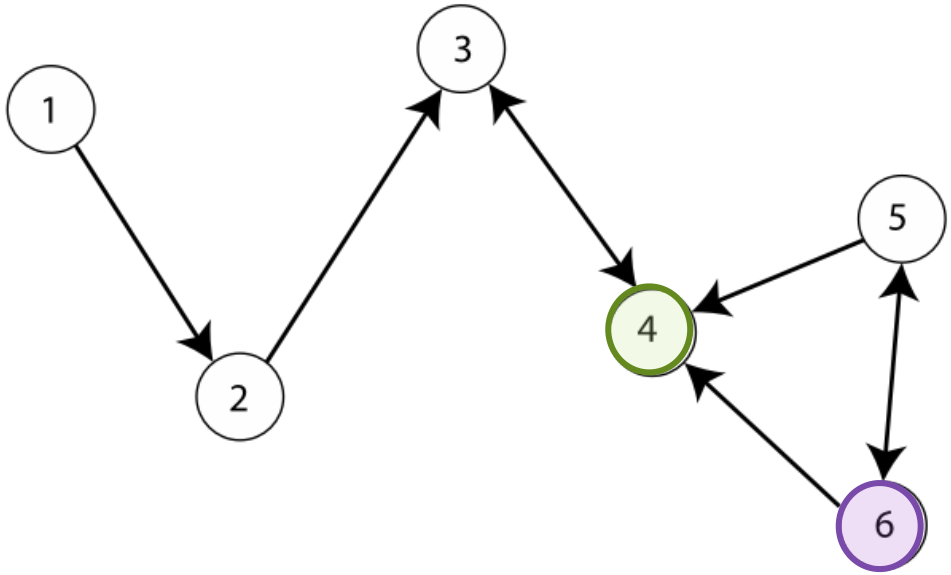
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	1	0	1
0	0	0	1	1	0



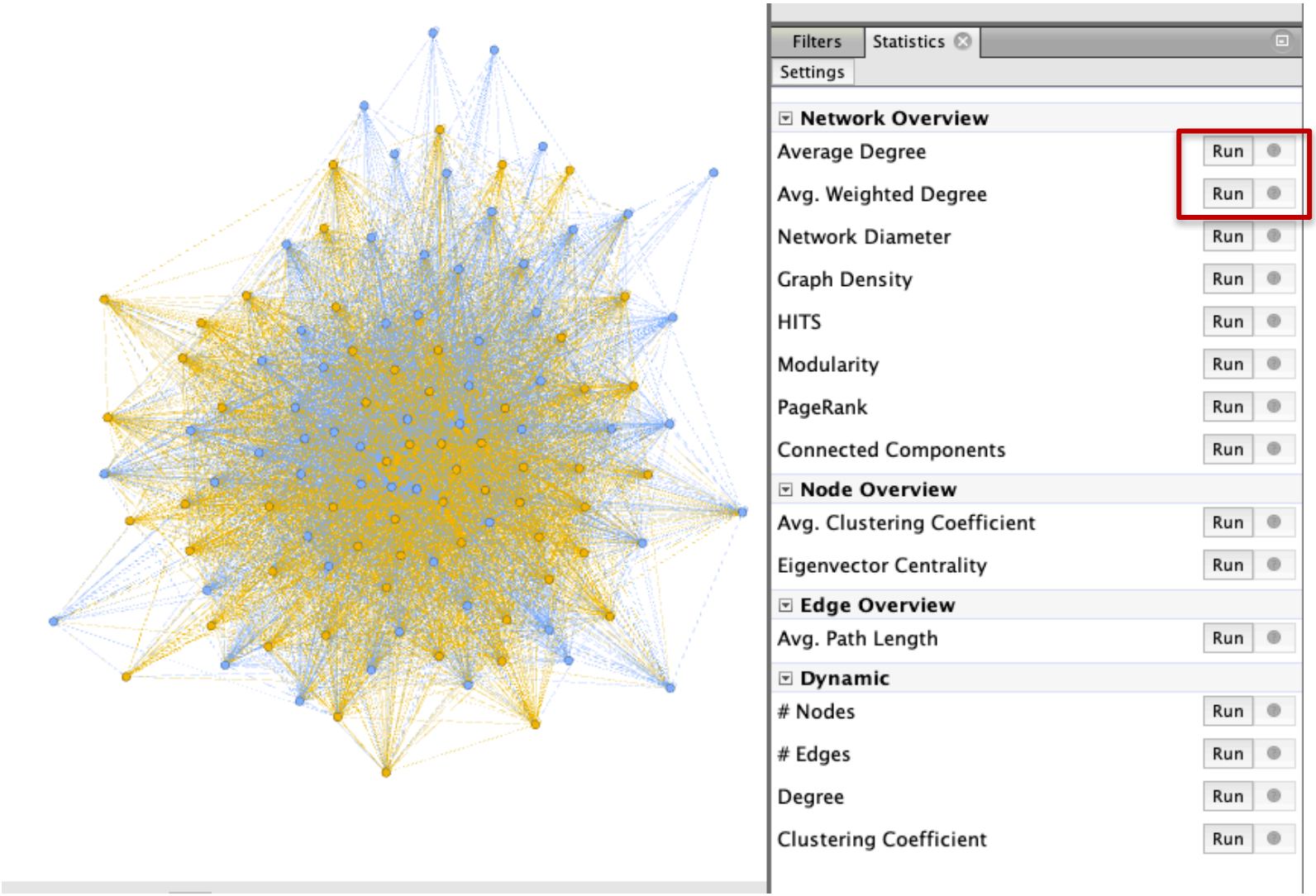
Degree

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

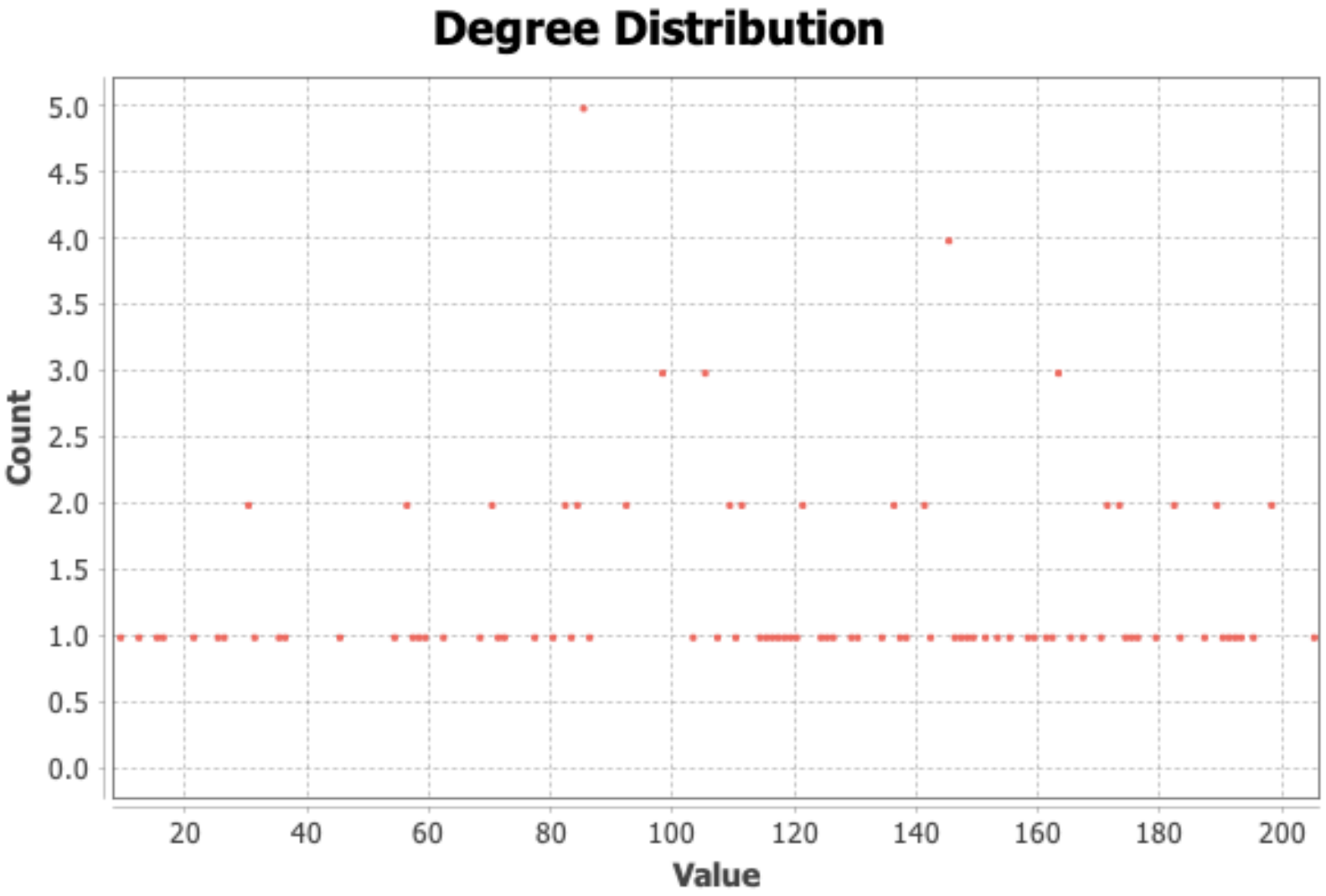
0	1	0	0	0	0
0	0	1	0	0	0
0	0	0	1	0	0
0	0	1	0	0	0
0	0	0	1	0	1
0	0	0	1	1	0



In Gephi…


Filters
Statistics
Settings
Network Overview
Average Degree
Run
Avg. Weighted Degree
Run
Network Diameter
Run
Graph Density
Run
HITS
Run
Modularity
Run
PageRank
Run
Connected Components
Run
Node Overview
Avg. Clustering Coefficient
Run
Eigenvector Centrality
Run
Edge Overview
Avg. Path Length
Run
Dynamic
Nodes
Run
Edges
Run
Degree
Run
Clustering Coefficient
Run

Worst. Visuals. Ever. 



(Yes, that should be a histogram.)

MSA Degree Comparison

MSA 2020

Average Degree: 57.990

Average Weighted Degree: 375.210

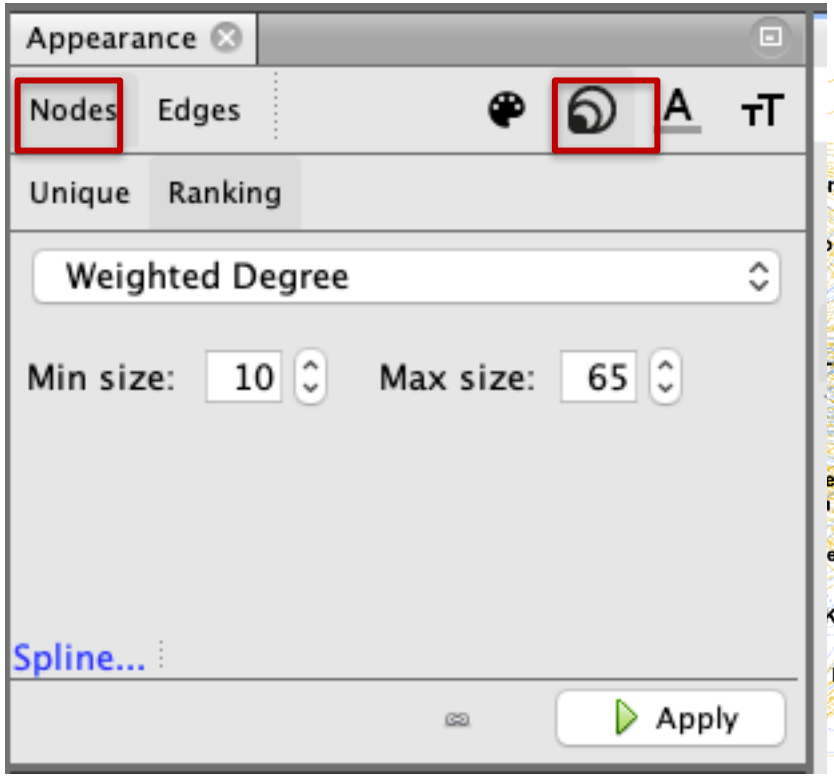
MSA 2021

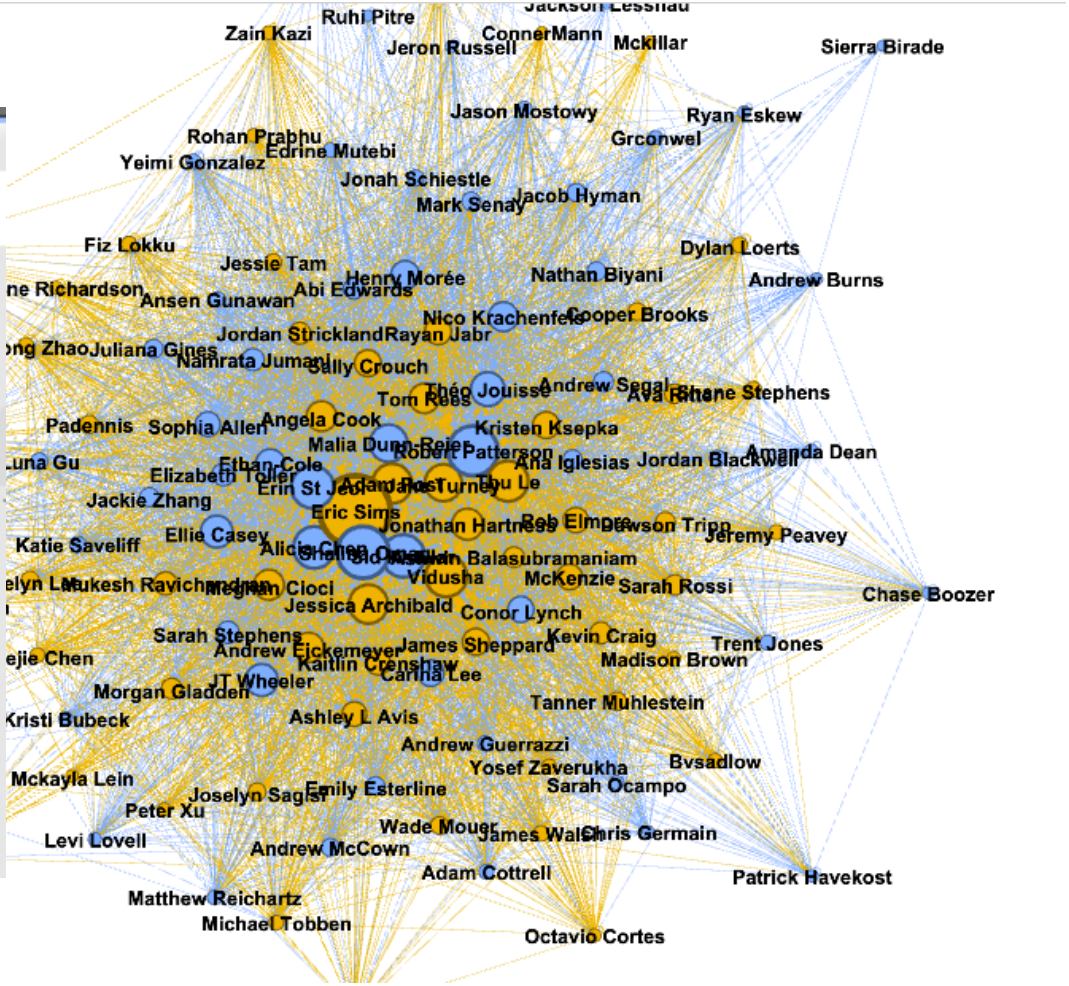
Average Degree:

Average Weighted Degree:

In Gephi...

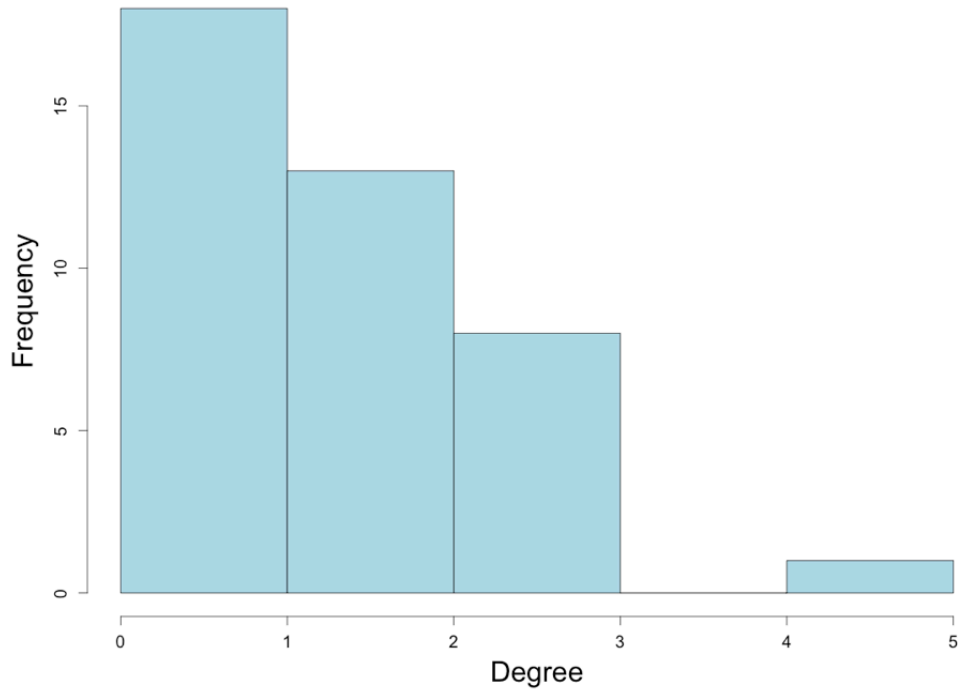
Any time you run a procedure from the statistics panel that computes a statistic, you'll see that information populate in Data Laboratory, which means you can use it in your viz.





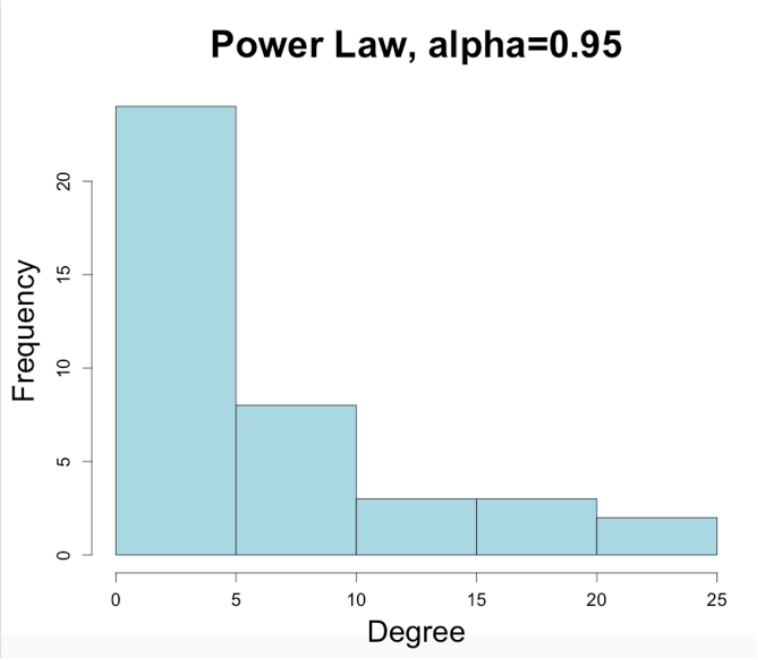
Degree Distribution

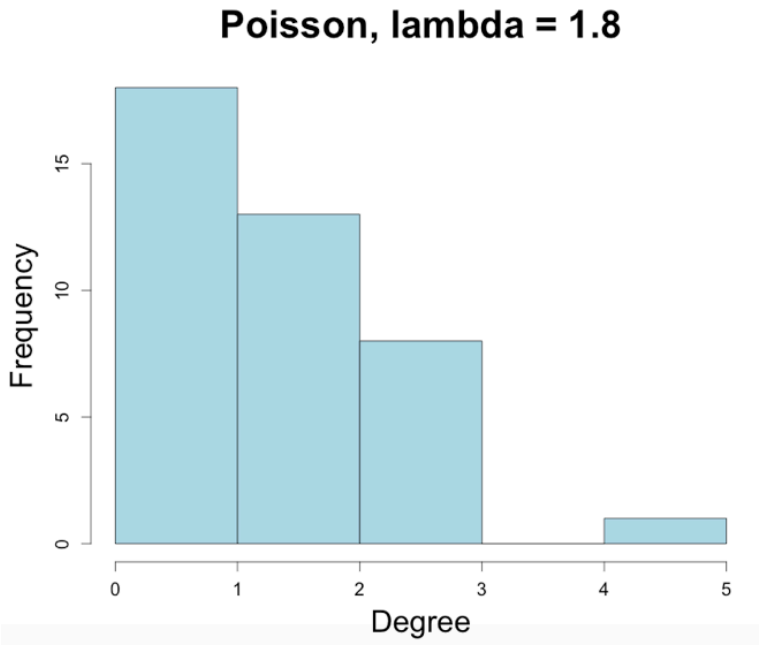
- It's common to look at the distribution of degrees in a network.
- Usually many nodes with low degree and few with high degree.



Degree Distributions

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



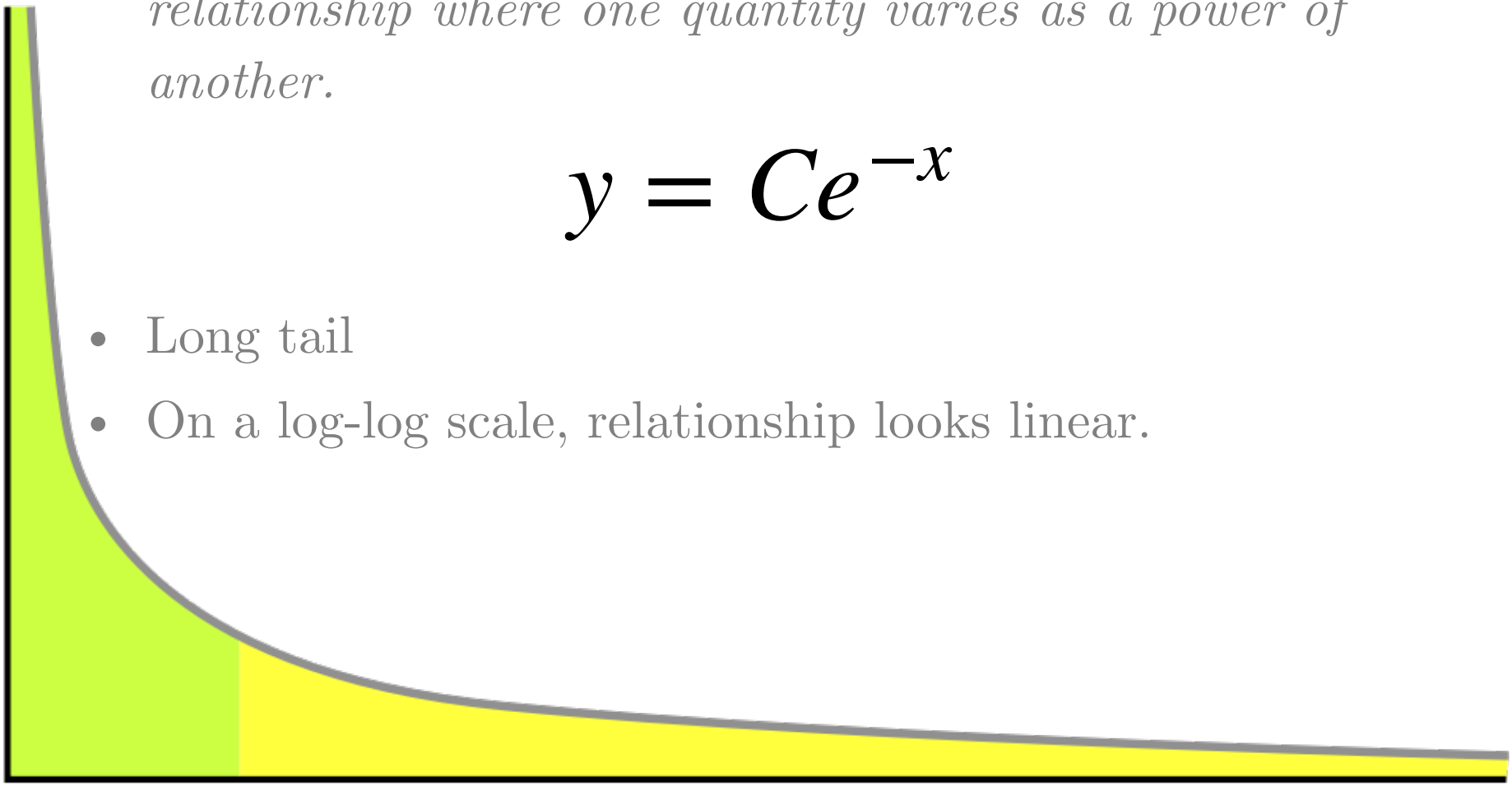


Power Law

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

$$y = Ce^{-x}$$

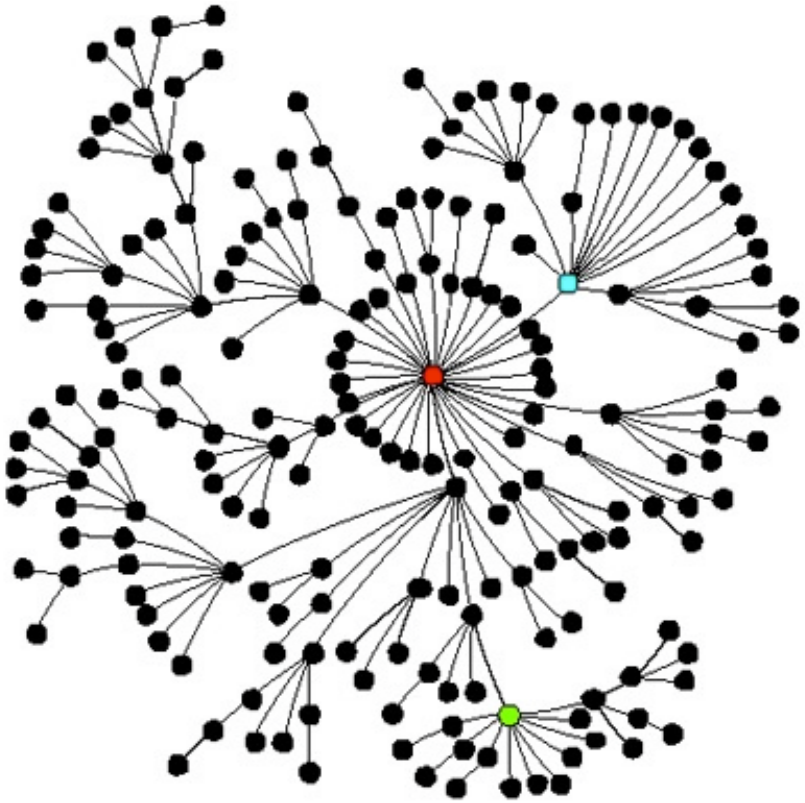
- Long tail
- On a log-log scale, relationship looks linear.



Power Law Graphs

aka Scale Free Networks

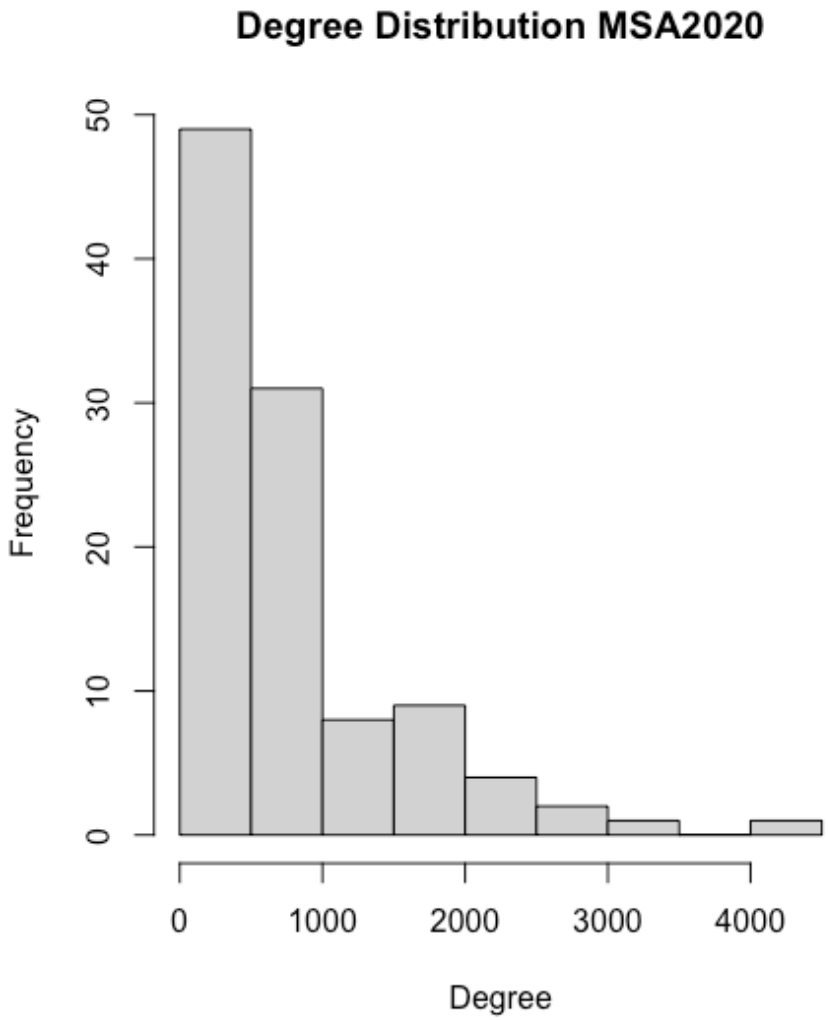
- Power law graphs contain a few **hubs** (highly connected nodes) but the majority of nodes in the network have low degree.

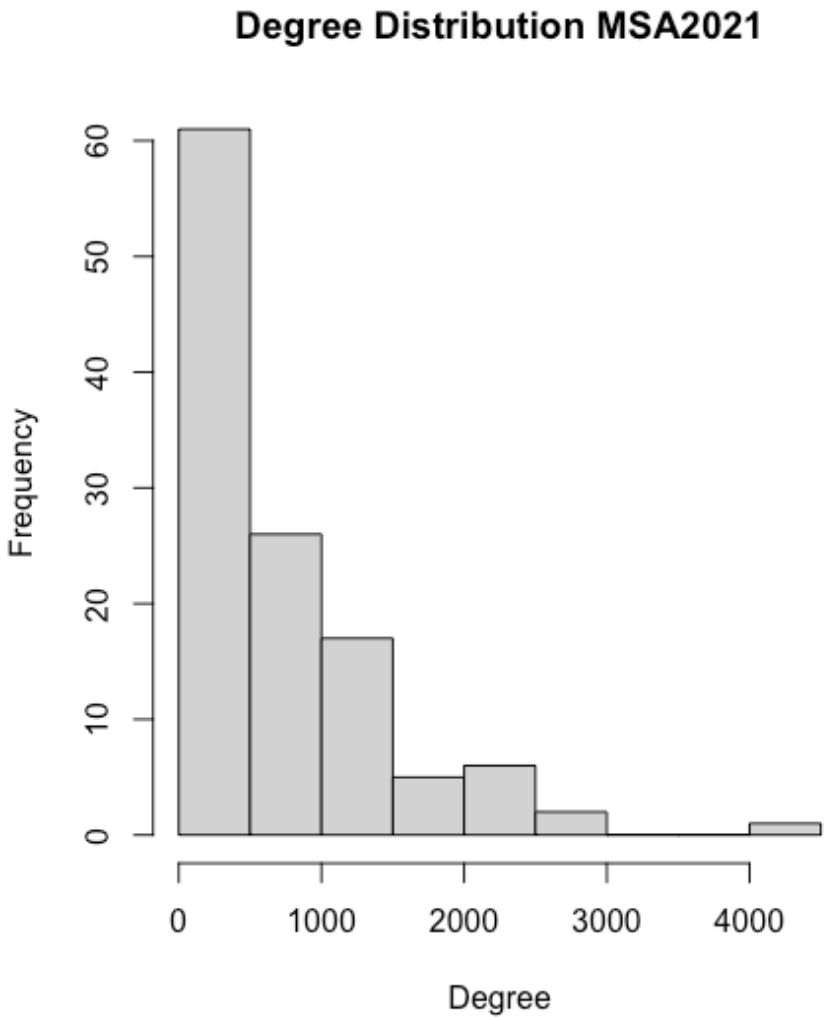


Power Law Graphs aka Scale Free Networks

- 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

MSA Degree Comparison





In R: `hist(degree(slack))`

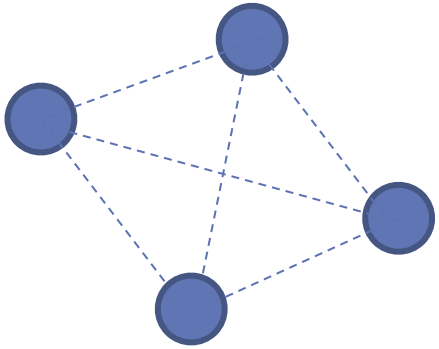
Other Descriptives

• • •

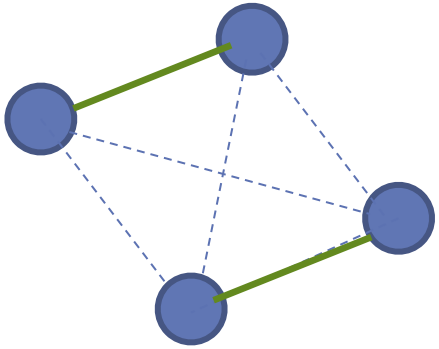
Density, Shortest Paths, Eccentricity, Clustering Coefficients

Density of Graph

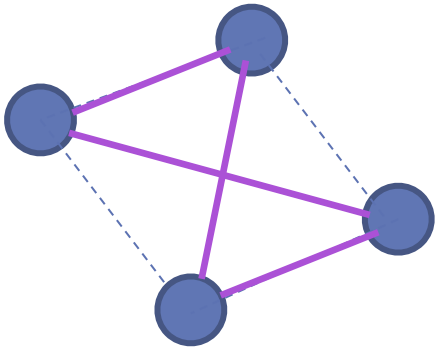
- The **density** of a graph measures how interconnected the nodes are.
- Simply the proportion of possible edges that actually exist in the graph.



6 possible edges



Density = 2/6 = 33%

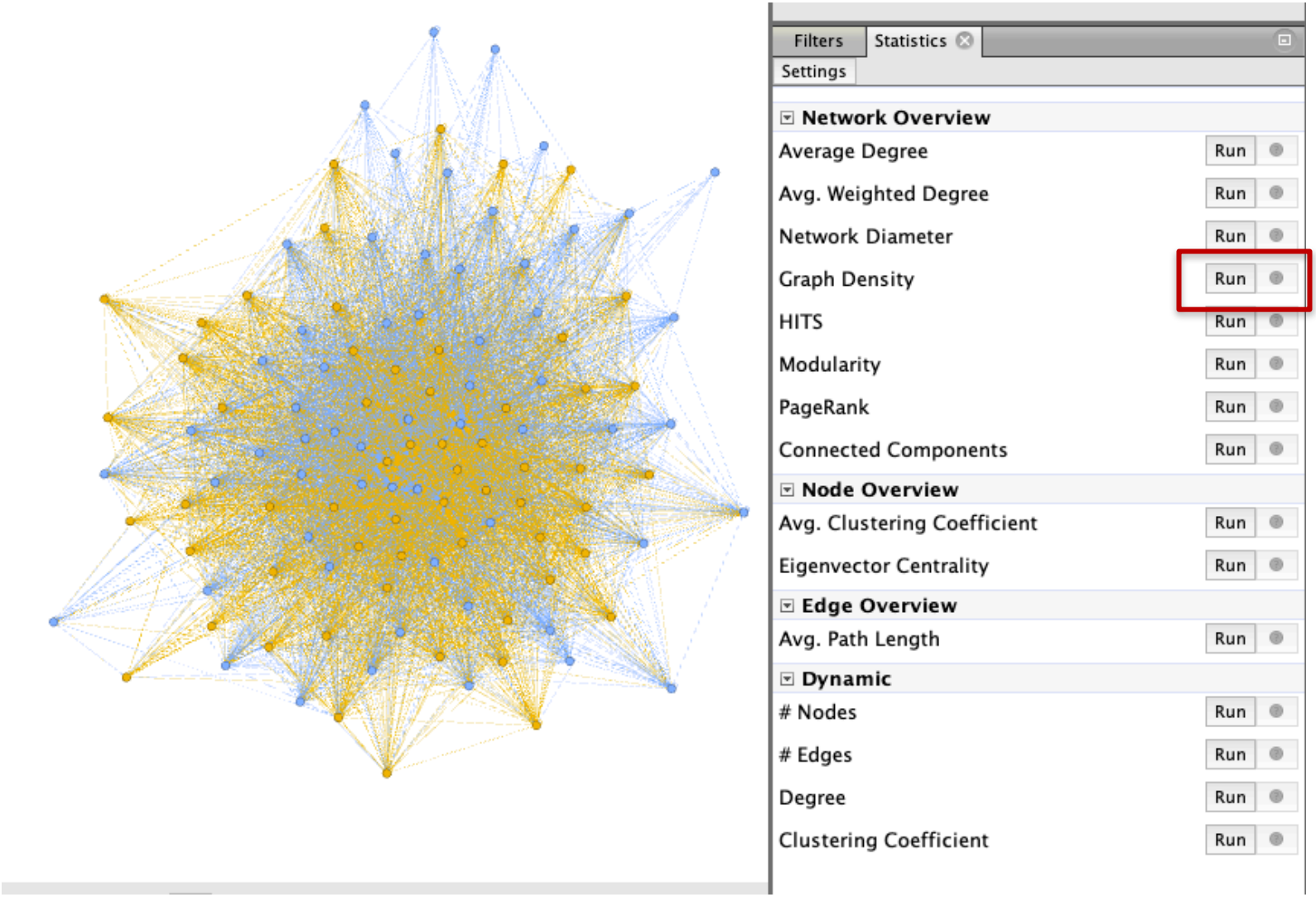


Density = 4/6 = 66%

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

In Gephi…


Filters
Statistics
Settings
Network Overview
Average Degree
Run
Avg. Weighted Degree
Run
Network Diameter
Run
Graph Density
Run
HITS
Run
Modularity
Run
PageRank
Run
Connected Components
Run
Node Overview
Avg. Clustering Coefficient
Run
Eigenvector Centrality
Run
Edge Overview
Avg. Path Length
Run
Dynamic
Nodes
Run
Edges
Run
Degree
Run
Clustering Coefficient
Run

MSA Density Comparison

MSA 2020

Directed Density: 0.558

Undirected Density: 0.711

MSA 2021

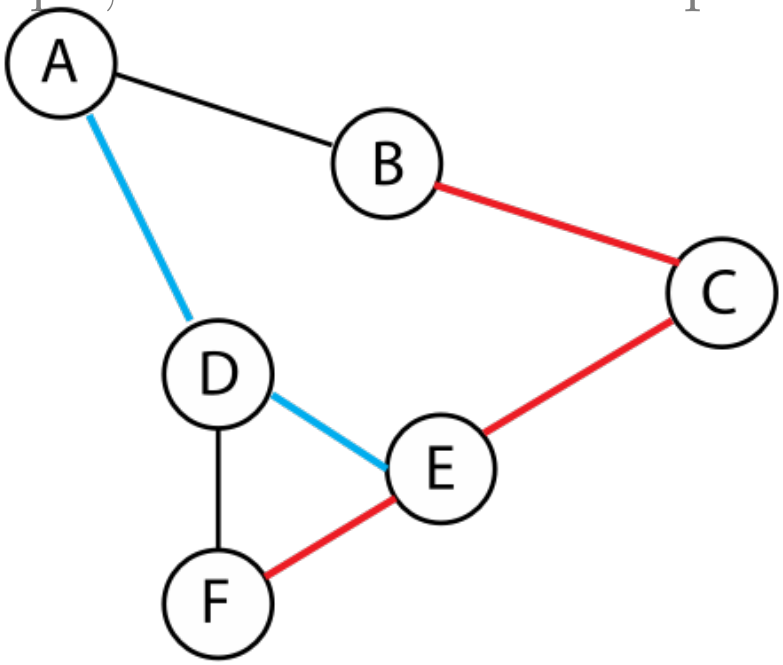
Directed Density:

Undirected Density:

Shortest Paths

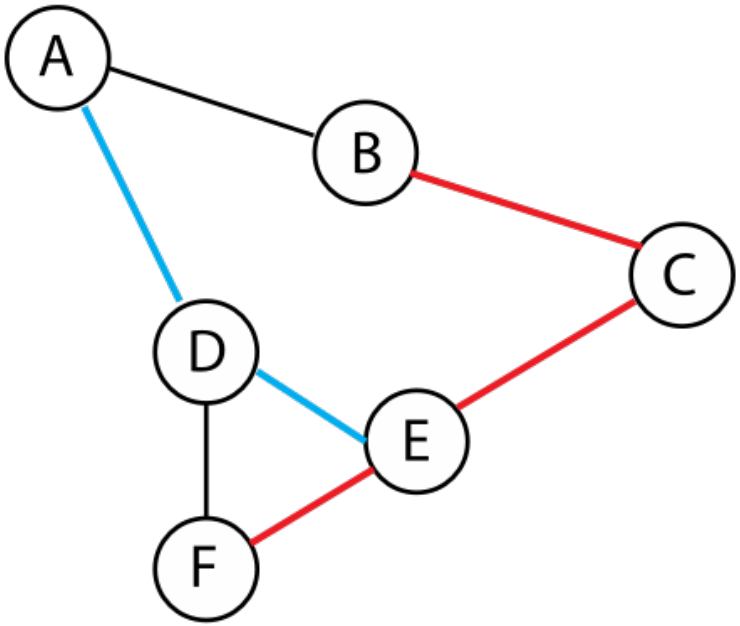
(Geodesic Distances)

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

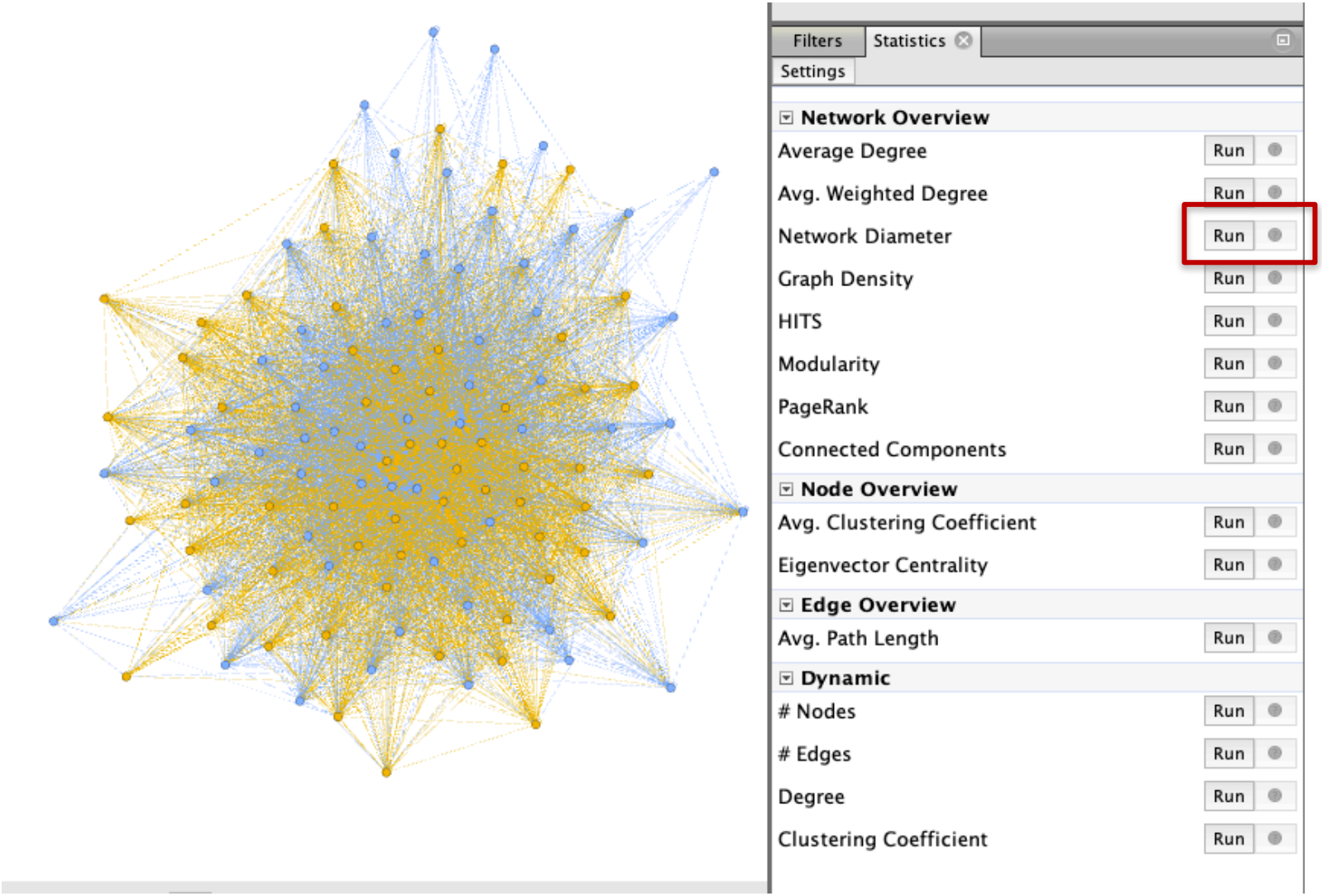


Diameter/Eccentricity

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



In Gephi…


Filters
Statistics
Settings
Network Overview
Average Degree
Run
Avg. Weighted Degree
Run
Network Diameter
Run
Graph Density
Run
HITS
Run
Modularity
Run
PageRank
Run
Connected Components
Run
Node Overview
Avg. Clustering Coefficient
Run
Eigenvector Centrality
Run
Edge Overview
Avg. Path Length
Run
Dynamic
Nodes
Run
Edges
Run
Degree
Run
Clustering Coefficient
Run

MSA Diameter Comparison

MSA 2020

Directed Diameter: 4

Average Path length: 1.445

MSA 2021

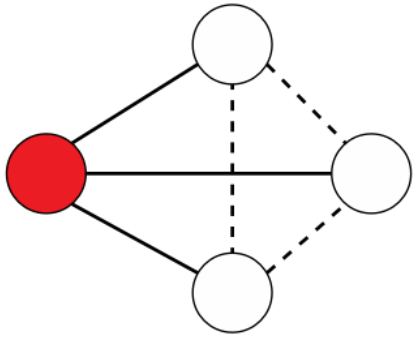
Directed Diameter:

Average Path length:

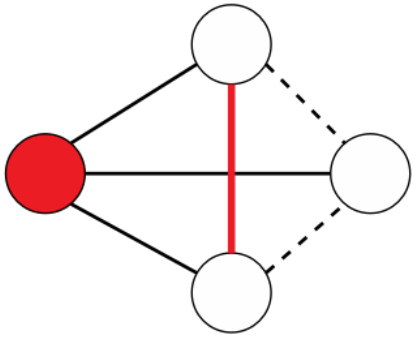
Clustering Coefficient

- The **clustering coefficient of a node** is a measure of the extent to which its neighbors are also neighbors of each other.
- It is the ***density*** *of the subgraph induced by the vertex.* (= that vertex and all of its neighbors and any edges between them)
- 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.

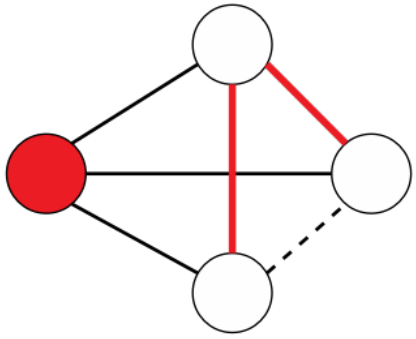
Clustering Coefficient



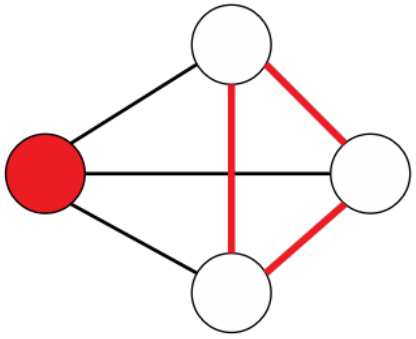
c=0



c=1/3



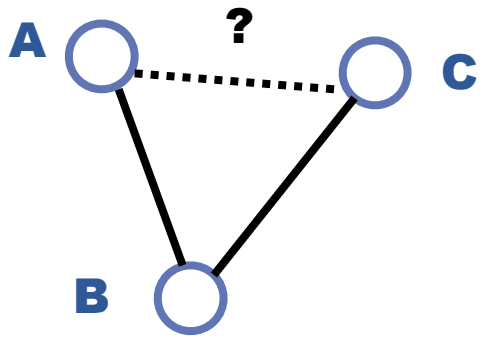
c=2/3



c=1

Clustering Coefficients for Entire Network

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



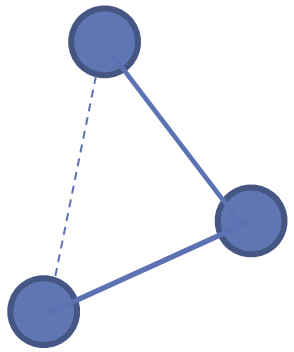
Clustering Coefficients for Entire Network

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

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

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

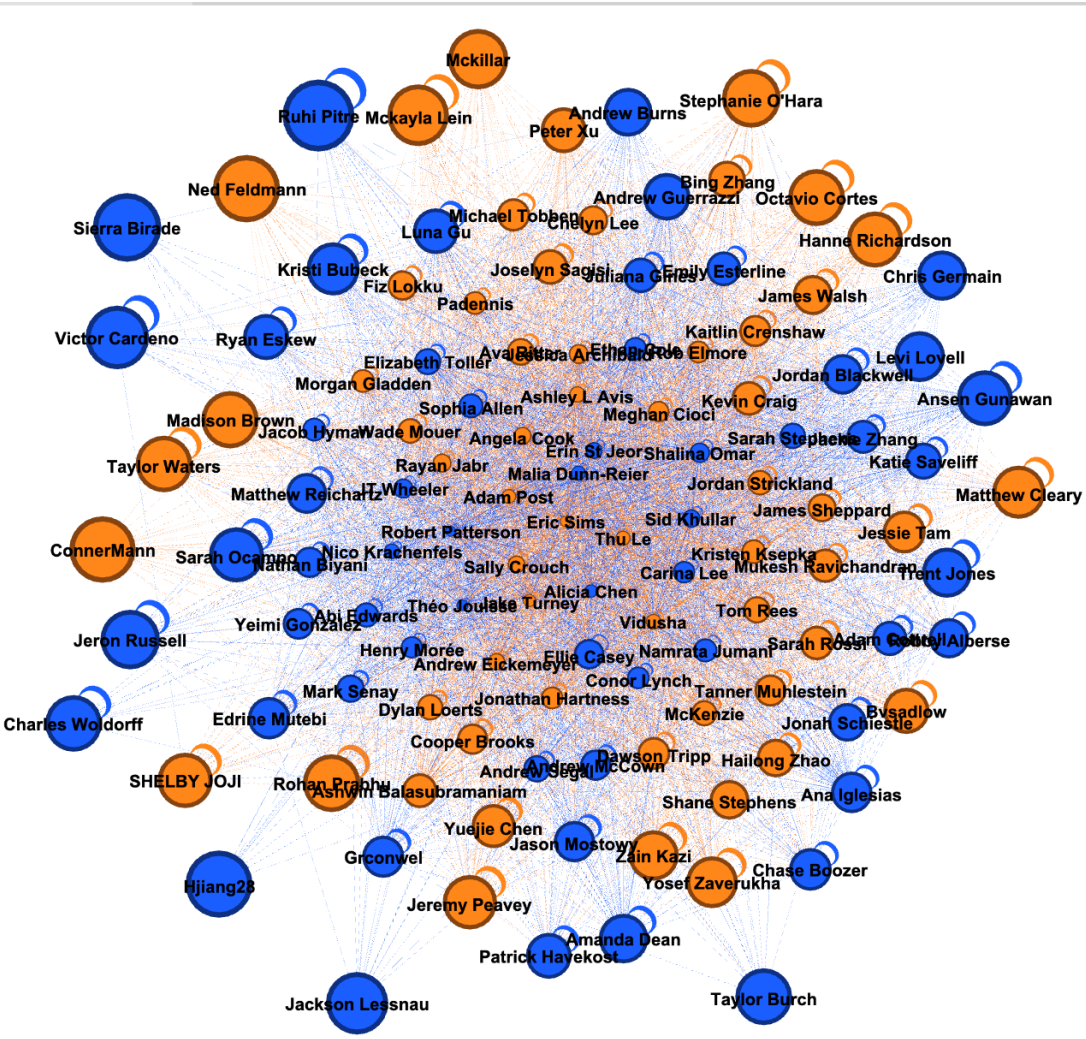
  - Connected triple is 3 vertices joined by 2 edges.
  - Each triangle makes 3 connected triples.



Utility of Clustering Coefficients

- "The clustering coefficient of a word [in a phonological network] has been shown to influence a number of language- and memory-related processes."
  —Goldstein, Vitevitch. U Kansas. 2017
- Directed clustering coefficient as a measure of systemic risk in complex banking networks.
- Relationship to success of cooperation in networks
- Clustering and preferential attachment in growing networks
- Clustering and the synchronization of oscillator networks

In Gephi…



Edges: 6923

Directed Graph

Filters | Statistics | Settings

Network Overview

Average Degree	Run
Avg. Weighted Degree	Run
Network Diameter	Run
Graph Density	Run
HITS	Run
Modularity	Run
PageRank	Run
Connected Components	Run

Node Overview

Avg. Clustering Coefficient	0.674	Run
Eigenvector Centrality		Run

Edge Overview

Avg. Path Length	Run

Dynamic

# Nodes	Run
# Edges	Run
Degree	Run
Clustering Coefficient	Run

MSA Clustering Coefficient Comparison

MSA 2020

Average Clustering Coefficient: 0.716

MSA 2021

Average Clustering Coefficient: